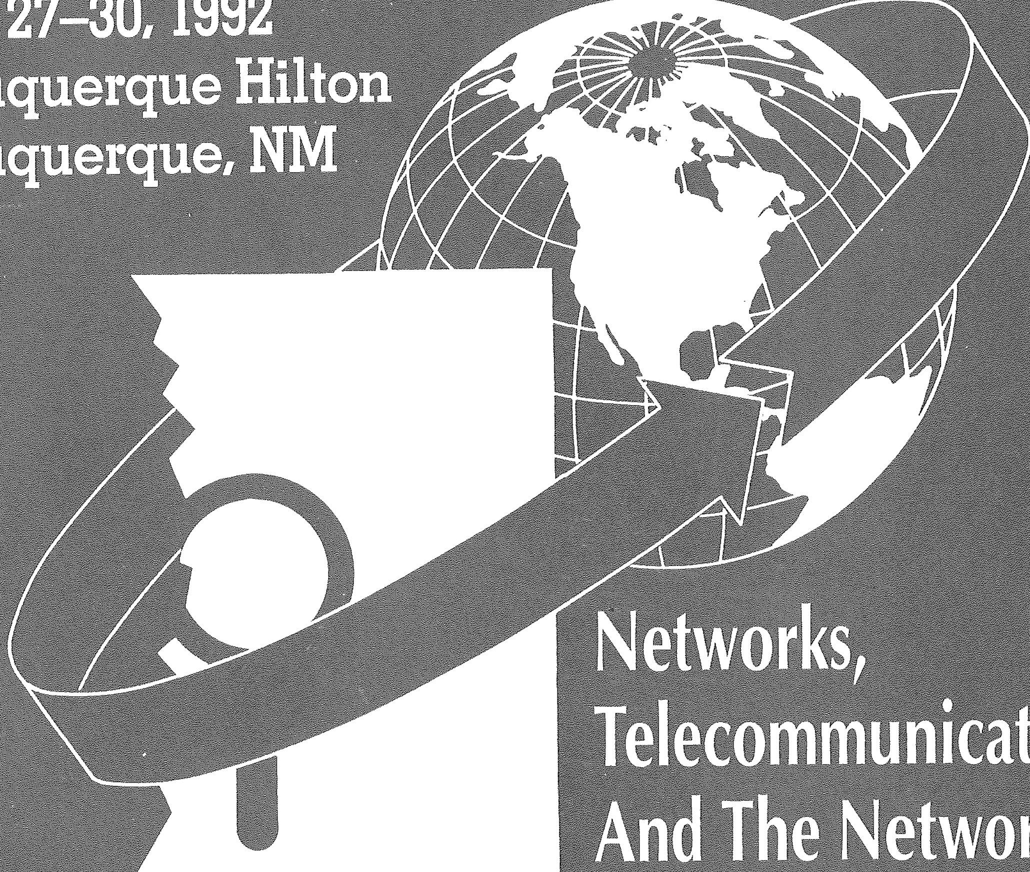
Proceedings ASIS Mid-Year Meeting May, 1992

asis

ASIS 1992 Mid-Year Meeting

May 27–30, 1992
Albuquerque Hilton
Albuquerque, NM



Networks, Telecommunications, And The Networked Information Resource Revolution

Proceedings

American Society for Information Science
8720 Georgia Avenue, Suite 501
Silver Spring, MD 20910
(301) 495-0900

ASIS 1992 Mid-Year Meeting

Networking, Telecommunications, and the Networked Informaton Revolution

May 28-30, 1992
Albuquerque, New Mexico

Technical Program Committee
Clifford A. Lynch
Michael Buckland
Cecilia Preston
Janet Vratney-Watts
Nancy Gusack

Published by the
American Society for Information Science
Silver Spring, Maryland
Voice (301) 495-0900
FAX:(301) 495-0810
Internet: asis@cni.org

The opinions expressed by contributors to this publication do not necessarily reflect the position or official policy of the American Society for Information Science.

1992 ASIS Mid Year Meeting

WEDNESDAY, MAY 27

Governance

11:00am–1:00pm
Int'l Relations Comm
N M Parlor B

Professionalism Comm
Jamez Room

12:00 – 1:30
Publications Comm
N M Parlor A

1:00pm – 3:30pm
Chapter Assembly
N M Parlor S

3:00pm–5:00pm
SIG Cabinet
N M Parlor S

5:00–6:30 pm
Standards Committee
N M Parlor A

Pre Conference Education Courses
9:00 am – 5:00 pm

TCP/IP and the Internet: A Tutorial
John Romkey
New Mexico North

Optical Disk Storage Technology & Appl.
James Rush
Nevada Room

Creating an Electronic Library: Building Blocks and Toolkit
David Batty
Arizona Room

Online Database Implementation: Taking a Database From the Idea Stage to a Working System
Marjorie Hlava
Jay Ven Eman
Texas Room

6:00pm–8:00pm — Welcome Reception — Arizona/Nevada Rooms

THURSDAY, MAY 28

	Arizona/Nevada Room	Colorado/Texas Room	New Mexico N., E&F
8:00am–9:00am	PC Telecommunications Software: A Tutorial **Bob Baker**	Soviet & E. European Resources Available on Elect Commun Svcs **Markiw**	

9:30am–11:00am **PLENARY SESSION** New Mexico, North & South, E –– H Rooms
The Internet Explosion
VINTON CERF, Corporation for National Research Initiatives

	Arizona/Nevada Room	Colorado/Texas Room	New Mexico N., E&F
11:30am–12:30pm	Fed. Natural Resourc & Energy Information Management Programs **Young, Splendoria Krumm**	Directories to Networked Info Contributed Papers **Moen; Guenther**	Libraries and National Networks **Lavagnino, Peters Parkhurst, Seiden**

12:30pm–2:00pm LUNCH BREAK

Awards and Honors Comm	New Mexico Parlor A
Membership Comm	Jamez Room
CISCO Comm	New Mexico Parlor B
Education Comm	Laguna Room

	Arizona/Nevada Room	Colorado/Texas Room	New Mexico N., E&F
2:00pm–3:30pm	State and Regional Networks **Lynch, Bajzek Shaw, Ogden**	Multimedia Issues In Networks **Kaye, Conley, Chuang Barb. Baker, Burnett Lesnansky**	Library Roles **Vratney–Watts McClure, Ryan, Lauterbach; Starnes, Sedayo Hambridge, St. Pierre**
4:00pm–5:30pm	Telecommuting in an Information Environment **Gresehover Sawyer, Milstead**	Info Retrieval: Contributed Papers **Janes & Rosenfeld Shepherd & Watters Tsai**	Spread Spectrum: Develop. of CAL State Packet Radio **Brownrigg**

5:30pm–7:30pm POOL PARTY RECEPTION

FRIDAY, MAY 29

	Arizona/Nevada Room	Colorado/Texas Room	New Mexico N., E&F
8:00am–9:00am	Economic Aspects Of Networks **T. Martin**	Doctoral Forum **T. Bearman**	Collection Develop. & the Elect. Library **Butler, Goodram Beatty, Kaufman Greenfield**

9:30am–11:00am **PLENARY SESSION** New Mexico, North & South, E –– H Rooms
The Independent Learner is an Empowered Citizen
Paul Evan Peters, Coalition for Networked Information

	Arizona/Nevada Room	Colorado/Texas Room	New Mexico N., E&F
11:30am–12:30pm	Electronic Communication Contributed Papers **Warshawsky; Peek**	Networked CD–ROM **Hudson; Gey**	Wide Area Information Servers: Special Session **Kahle**

12:30pm–2:00pm LUNCH

Research Comm	New Mexico Parlor B
Public Affairs Comm	Laguna Room
Conf. & Meetings Comm	Jamez Room
Const. & Bylaws Comm	New Mexico Parlor A

	Arizona/Nevada Room	Colorado/Texas Room	New Mexico N., E&F
2:00pm–3:30pm	Changes in Communication Patterns **Steele; Liebscher; Gregory**	Online Catalogues And Networks **Buckland, Norgard, Plaunt; Dillon,Jul; Jeng**	Large Scale Networked Information Projects: Special Report **Lynch, Larson, Saylor**
4:00pm–5:30pm	Progress Towards Remote Image Serving: Case Studies in A & H **Busch, Egan, Giral Dackow**	Information on The Internet: Discovery Access & Use **Dillon, Kahle, Kovacs Schwartz, Seiden**	

6:00pm–9:00pm PUEBLO CULTURAL CENTER FEAST (meet in lobby)

SATURDAY, MAY 30

	Arizona/Nevada Room	Colorado/Texas Room	New Mexico N., E&F
8:00am–9:00am	Library Automation & Networking: Issues & Opportunities **Doty, Hert**	A Coop File–Library for Elect. Exper. The Consortium for Lexical Research **Wilks, de Ram**	

9:30am–11:00am **PLENARY SESSION** New Mexico, North & South, E –– H Rooms
Libraries Made with Silicon and Fiber Optics
Brewster Kahle, Thinking Machines Corporation

	Arizona/Nevada Room	Colorado/Texas Room	New Mexico N., E&F
11:30am–12:30pm	User Studies & Eval. in Net Environments **Bishop, Doty, McClure; Murray; Raman**	Bio & Chem Info In Network Environs **Hurd, Llull Bishop, Liebscher**	Z39.50: Protocol and Reality **Lynch, Needleman M. Baker**

12:30pm–2:00pm LUNCH

	Arizona/Nevada Room	Colorado/Texas Room	New Mexico N., E&F
2:00pm–3:30pm	Benchmarking in Labs and Corporations **Allan, Coers, Byrne Longmire, Trujillo**	Advanced Computer & Engin. Research To Serve Med. Info. **Thoma, Ackerman Sheretz, Benson**	Educating the Network Info Professional **Shaw, Barker, Newby Verkade, Doctor**
4:00pm–5:30pm	NREN Update	Advanced Computer & Engin. Research To Serve Med. Info. (Above)	

6:00 pm – 9:00 pm **ASIS Board of Directors** New Mexico Parlor A & B

CONTENTS

Part I: Contributed Papers

Part II. Panel Sessions

Author Index

Part I

Contributed Papers

Soviet and East European Studies Information Resources Available on Scholarly Electronic Communications Networks

Michael Markiw
Arizona State University Libraries, Tempe, Arizona

ABSTRACT

The recent proliferation of scholarly electronic communications networks has lead to a growth of networked information resources dealing with specialized areas of interest. This paper focuses on the variety of information relating to Soviet and East European studies which is available on academic communications networks. Within networks such as BITNET and Internet access is available to information resources such as online library catalogs, online data bases, listservers supporting conferences of special interest groups, bulletin boards, electronic journals and newspapers, newsletters, and files transferred through FTP. Networked resources of this type which may be of particular interest to Soviet and East European scholars will be discussed along with methods of access to them.

Electronic mail might not be strictly regarded as an information resource but one must consider its potential for enhancing scholarly communication and providing timely and valuable information within the field of Soviet and East European studies. Scholars are now able to obtain primary information directly from Soviet citizens and groups through telecommunications links with some Soviet computer networks. A description of such links within this paper might help to provide the opportunity for scholars to establish or to increase communication with their counterparts in the Soviet Union.

INTRODUCTION

For many years scholarly information resources were limited to monographs, serials and other information appearing in mostly paper and film formats. However, the advent of electronic communications networks has led to a recent proliferation of information available in electronic form on these networks. In 1986 Oberst noted that "the past five years have seen dramatic growth in the use of networking for scholarly and administrative communication, as a result of proliferation in the number and types of computer networks at every level: departmental, campus wide, regional, national and international...BITNET has enabled academics from virtually every field to experience the value of electronic communication."[1] Two of the national networks, BITNET and the Internet, have become major contributors to this recent dramatic growth in scholarly communications. In 1990 Arms found that BITNET and the Internet "are already essential tools for many researchers, providing access to a growing array of information sources."[2]

Within BITNET and Internet, access is available to information resources such as online library catalogs, online databases, listservers supporting conferences of special interest groups, bulletin boards, electronic journals and newspapers, newsletters, and files transferred through FTP. The focus of this paper will be on describing some of these network information resources in the field of Soviet and East European studies.

DISCUSSION

Certain online library catalogs are available on the Internet and are accessible without charge by using the TELNET command followed by the network address in order to establish a connection. For example, to access the University of Michigan's library catalog enter TELNET cts.merit.edu or TELNET 35.1.48.150. For more information contact Dave_Katz@um.cc.umich.edu. A scholar may be interested in searching the catalogs of research libraries which often have large collections on the Soviet Union and Eastern Europe and may hold strong collections on certain historical or literary figures and periods important to Soviet and East European studies. Some research libraries with catalogs on the Internet are University of California at Berkeley, University of Kansas, University of Michigan, University of Minnesota, University of Texas, Columbia University, Ohio State University, Princeton University and New York Public Library. A list of online library catalogs accessible through the Internet can be obtained by sending the command GET INTERNET LIBRARY to listserv at unmvm.

Some online databases available on the Internet and accessible for charge are RLIN (Research Libraries Information Network) and OCLC's Epic Service. RLIN offers its online catalog together with flexible searching procedures which allow for subject and keyword searches as well as the ability to limit by boolean operators. However, Internet access does not permit Cyrillic displays which have been available since 1986 on dedicated terminals because of a required software package installation. For RLIN subscription information contact Martha Girard at the Internet address bl.mxg@rlg. Among the services which Epic offers is a database of the OCLC Online Union Catalog which provides a series of search indexes of records for books, serials and other materials. Searching by subject, keyword and boolean limiters is also available. Of particular interest for Soviet scholars might be the subject search with a language restrictor. This means that all titles on a particular subject and published in a particular language would be retrieved, e.g. all Russian language titles on Russian poetry. In its Epic Service news messages OCLC advises those who are not sure whether their institution has Internet access to contact the institution's computing center regarding Internet account information. For Epic service subscribers the Internet addresses are 132.174.100.2 or epic.prod.oclc.org.

Computer conferences of special interest groups are accessible on both BITNET and the Internet. According to Britten "perhaps the most practical and worthwhile use of BITNET is the interest group lists. There are hundreds of discussion groups active on the networks."[3] In 1991 Marker notes that "there are 900 computer conferences on Bitnet. Typically these conferences are supported by list servers, and they are called 'lists.'"[4] Some conferences on BITNET and other networks which should be of interest to Soviet and East European scholars are talk.politics.soviet, soc.culture soviet, RUSSIA (Russia and Her Neighbors List), VAL-L (Valentine Michael Smith's Commentary), BALT-L (Baltic Republics Discussion List), soc.culture.magyar, soc.culture.polish, soc.culture.yugoslavian, SEELANGS (Slavic and East European Languages and Literature List), RUSSIAN (Russian Language Issues), and RUSTEX-L (Russian Tex and Cyrillic Processing List).

Talk.politics.soviet, as the name implies, focuses on Soviet political issues. During the coup of August 19-21, 1991 this group was instrumental in providing to the rest of the world current reports on the situation. Russian citizens' descriptions of events as they were occurring were transmitted over this computer network together with reports directly from various Soviet news agencies. Reports also included official statements by Soviet and Russian officials. Among the most important was Russian President Boris Yeltsin's decree of August 19, 1991 stating that the president of the USSR (Mikhail Gorbachev) was dismissed in a coup attempt and this is considered a state crime so the government agencies of the RSFSR are to execute the functions of the corresponding bodies of the USSR and to prevent execution of orders from the unconstitutional coup committee. Versions of this decree were available in English and Russian although the Russian form was a transliteration from the Cyrillic alphabet. Some post-coup topics include economic union among the former USSR republics, why republics such as Armenia, Georgia and the Ukraine should or should not become partners in the new economic union, the future of the Baltics, and current events in the Soviet Union. Occasionally news bulletins issued by Soviet and international press agencies are received. Talk.politics.soviet is available as a newsgroup on Usenet, a large international computer network, or subscriptions can be requested on BITNET from listserv@indycms under the list name TPS-L.

Soc.culture.soviet deals with Soviet cultural issues. Some typical subjects discussed are Slavic customs, political and other jokes, how to mail money and other items to the USSR, translation techniques, Russian and other Soviet literatures, religion, history of Russia and other republics, and cooking. During the coup of August 19-21, 1991 this network's attention was also directed toward reports and discussions of the situation and many of the same items appeared simultaneously on talk.politics.soviet. Soc.culture.soviet is also available as a newsgroup on Usenet member sites or can be subscribed to on BITNET via listserv@indycms where it is known as SCS-L.

RUSSIA (Russia and Her Neighbors List) is concerned with the new order within the new Soviet Union and political affairs of neighboring countries. Along with the two networks mentioned above this group was also a medium for transmitting information on coup events as they were occurring. Some recent topics include memoirs from the coup, post-coup analyses of the economic union among Russia and other former republics, and the problem of naming this new post-coup political entity. Potential foreign relation between this new state and surrounding countries such as the Baltics and Yugoslavia is a strong area of interest. This network has become a source for information on the current political situation within countries bordering on the new Soviet Union. For example, Croatian unrest within Yugoslavia is monitored. Subscriptions to RUSSIA can be placed through BITNET on listsev@indycms on CREN and listserv@indycms.iupui.edu on the Internet. At this time it is not available on Usenet.

VAL-L (Michael Valentine Smith's Commentary) offers opinions by list owner Valentine M. Smith on the changing state of the communist countries. In addition to being provided with Mr. Smith's commentaries, list members share their views in often article-length essays which may discuss in great detail political and social issues within the new Soviet Union and adjacent countries as well as in other communist governed states. Occasionally the exchange of viewpoints extends to domestic and international issues not necessarily related to communist countries. VAL-L is available on BITNET from listserv@ucf1vm. Some other lists which provide opportunities for more general political discussion and occasionally include Soviet and East European affairs appear as newsgroups on Usenet under the names alt.activism, alt.activism.d, alt.conspiracy, misc.activism.progressive, misc.headlines and soc.rights.human.

BALT-L (Baltic Republics Discussion List) discusses politics, current affairs and provides general information on Estonia, Latvia and Lithuania. Topics have included relations among the Baltic republics and neighboring countries, Baltic citizenship, political leaders, economic issues, Baltic Americans, treatment of ethnic minorities in the Baltics, and telecommunications links to this area. This list is available from listserv@ubvm on BITNET.

Some discussion groups dealing with specific Eastern European countries are soc.culture.magyar, soc.culture.polish and soc.culture.yugoslavian. Hungarian, Polish and Yugoslav culture and politics are the main interests and contributions are occasionally offered in the languages of these countries. These three networks are available as newsgroups on Usenet member sites. Newsgroups on Bulgaria and Romania have been proposed.

SEELANGS (Slavic and East European Languages and Literatures List) is a vehicle for scholarly communication among members of the American Association of Teachers of Slavic and East European Languages (ATSEEL) but is open to non-members. Members might, for example, inquire where the next conference will be held or request

an e-mail address for the ATSEEL Newsletter so they could send submissions. General information on the organization is available and may include lists of officers and information on meetings and programs. This body also maintains a database of its Committee on College and Pre-College Russian which reports on committee meetings. Discussion topics are not limited to organizational activities. Soviet satellite television program schedules are regularly listed and requests were made for names of qualified appraisers of Slavic books in the Mid-Atlantic area and for sources of East European periodicals, especially Bulgarian newspapers. Subscriptions can be addressed on BITNET to listserv@cunyvm.

RUSSIAN (Russian Language Issues) is devoted to Russian linguistics, grammar, translation and literature. Submissions are preferred in Russian but English is acceptable. Since Cyrillic characters are not yet represented on network computer screens, Russian submissions are represented in a variety of transliteration schemes. The Library of Congress scheme is among the most popular perhaps due to its widespread use in U.S. academic libraries. Another popular scheme on this network is KDI-7 because it can be used to convert non-Cyrillic text to Cyrillic through installation of the appropriate hardware and software package. Other methods of representing Cyrillic characters on pc's are also discussed. Some other subjects addressed are Russian business letters, forms of address among Russians since the attempted coup, Russian phraseology, style, semantics, and criticism of Russian poetry. Subscriptions to RUSSIAN can be requested on BITNET from listserv@asuacad.

RUSTEX-L (Russian Tex and Cyrillic Processing List) is concerned with representation of Cyrillic characters on computer screens but a greater percentage of contributions is devoted to this subject than on the RUSSIAN list and the topics often are more technical. Some typical subjects are: use of Cyrillic text processing systems such as Russian Tex; transliteration of other Slavic Cyrillic alphabet languages such as Ukrainian, Serbian, Belorussian, Bulgarian and Macedonian; transliteration of non-Slavic Cyrillic alphabet languages such as Bashkir and of non-Slavic non-Cyrillic languages such as Armenian; which Cyrillic fonts to use with which printers for downloading; Cyrillic typefaces, and how Cyrillic software packages affect keyboard mapping. RUSTEX-L is available on BITNET from listserv@ubvm.

SUEARN-L (Connecting the USSR to Internet Digest) is an electronic journal which provides information on telecommunication links with Eastern Europe. Articles may deal with directions on reaching Soviet sites by electronic mail, how modems and other equipment work over Soviet phone lines, technology export restrictions, prospects for connecting more sites to Internet, the Soviet Union's online industry, what online services are available, Soviet user profiles, access of Soviet users to foreign databases, access of foreign users to Soviet databases, Soviet online contacts, communicating with Czechoslovakia, and the Soviet computer networks GlasNet and RELCOM. Subscriptions to SUEARN-L can be placed

through listserv@ubvm and it is available on Usenet member sites as newsgroup bit.listserv.su-earn. Back issues are available by anonymous FTP from impaqt1.mem.drexel.edu (129.25.10.1) in the /pub/suearn/back-issues directory, or from the list moderator, Michael Meystel, meystma@duvm.ocs.drexel.edu in e-mail or printed form.

Thus far this paper has focused on electronic communications networks in the West which provide information about Soviet and Eastern Europe. However, scholars interested in obtaining primary information directly from Soviet citizens and groups can explore direct electronic mail links with some Soviet computer networks. Since details of such links would require more time than permitted here this topic will be only briefly treated.

GlasNet, the newest computer network in the Soviet Union as of this writing, began operation in May 1991. It offers information exchange within the USSR among such diverse groups as scientists, educators, cultural groups, journalists and environmentalists. These groups are also able to enter into worldwide electronic communication with their counterparts. Information about GlasNet's electronic mail and conferencing services can be obtained from David Caulkins, San Francisco office director, through Internet at dcaulkins@igc.org or Anatoly Voronov, Moscow office director, through GlasNet at avoronov@glas.apc.org.

Relcom (Russian Electronic Communications), another recently established Soviet computer network, links universities, research institutions and government agencies in approximately 70 Soviet cities. During the August coup of 1991 Soviet resistance forces used this network to keep the rest of the world informed on the situation by providing news and even transcripts of Russian President Boris Yeltsin's speeches to the rest of the world. Vadim Antonov, a member of the Moscow software cooperative Demos and a Relcom founder, transmitted much of this information to the West. Archival files of his messages can be obtained from some of the above-mentioned discussion groups by sending the INDEX command to the listserv connected with that particular group and then scanning files sent from avg@kremvax.hq.demos.su.

Concerning other Soviet networks, as of 1990, according to Quarterman, "there have been few network connections to the Soviet Union .. the advent of glasnost seems to be having an effect on the situation, however."[5] He then lists some large Soviet networks and their connections with Eastern Europe. Since that time the two above-described networks GlasNet and Relcom have sprung up. For those wishing to inquire about e-mail addresses of firms and individuals in the Soviet Union, a list of Internet nodes within that country has been compiled by Fedor Pikus and is available as a file for retrieval from listserv@indycms (CREN) or listserv@indycms.iupui.edu (Internet). In addition, network services are being expanded to provide electronic information resources through subscription. For example, GlasNet now offers a weekly electronic version of <u>Moscow News</u> as well as a fax digest of

this periodical. More information can be obtained from mosnews@glas.aps.org.

CONCLUSION

Through a description of various electronic information resources concerning the Soviet Union and Eastern Europe, including electronic communications networks, it is hoped this paper will help make Soviet and East European scholars aware of these sources of often current information on their subject areas. At the same time, the section on electronic mail links may help to provide the opportunity for scholars to establish or to increase communication with that area of the world.

NOTES

(1) Daniel J. Oberst and Sheldon B. Smith, "BITNET: Past, Present, and Future," Educom Bulletin, (Summer 1986), 10.

(2) Caroline R. Arms, "Using the National Networks: BITNET and the Internet," ONLINE, (September 1990), 24.

(3) William A. Britten, "BITNET and the Internet: Scholarly Networks for Librarians," College and Research Library News, (February 1990), 105.

(4) Rhonda J. Marker, "Networks for Neophytes: An Introduction to Access and Use of Regional and National Telecommunications Networks," LITA Newsletter, (Winter 1991), 20.

(5) John S. Quarterman, The Matrix: Computer Networks and Conferencing Systems Worldwide (Bedford, MA: Digital Press, 1990), 506.

Organizing Networked Resources for Effective Use:
Classification and Other Issues in Developing Navigational Tools

William E. Moen
School of Information Studies
4-206 Center for Science and Technology
Syracuse University
Syracuse, NY 13244-4100
Phone: (315) 443-2911
Fax: (315) 443-5806
Internet Mail: WEMOEN@RODAN.ACS.SYR.EDU
BITNET Mail: WEMOEN@SUVM.BITNET

Introduction

Computer networks and the resources made accessible by these networks provide new opportunities for people to gather, use and share information. The networks range from local area networks with resources housed on local servers to large national electronic networks such as the Internet. The latter network, a logical network consisting of many interconnected local, wide-area, and backbone networks, offers a multitude of information and computing resources.

More networked resources and services become available daily, and it becomes more daunting to locate and use them. More computers are internetworked. Files that were once local may now become part of the larger networked resources universe. New electronic information is generated at ever-increasing rates as new discussion lists are established, more people post more messages to newsgroups, electronic journals are established, and other resources appear.

The nature of the electronic network environment and the networked resources and services available have special characteristics that librarians and information professionals need to understand. The resources themselves are dynamic and volatile. Contents of them change by the day, the hour, even the minute. The stability found in printed materials such as books and journals is absent. Networked resources will not necessarily be acquired, as in the case of traditional library materials. Access to these resources rather than ownership becomes more important. A major task for librarians and information professionals is to identify and enable access to them.

A primary barrier to effective use of existing networked resources and services is the lack of adequate locator systems, directories, guides, and indexes. To use resources available on the networks, network users need to identify, locate, and understand them. Navigational tools should assist users in these activities. Some navigational tools are available, but these are usually limited in scope, not comprehensive, and are themselves not fully identified or familiar to network users. Because the networks have developed in a piece-meal, decentralized manner, lack of comprehensive navigational tools is understandable.

This paper identifies some key issues and problems in developing adequate navigational tools to assist users in effective use of the networks. Specifically its focus is on how categorizing networked resources and developing appropriate classification structures can inform the intellectually organization of the resources and the development of navigational tools. The organization of information has been a vital activity of librarians and information professionals. The network environment offers a new opportunity to use the experience gained from bibliographic organization. We need to build upon what we know to

produce effective organizational schemes that fit the network environment. In addition, this paper asserts that network users and their information behaviors and needs can be a point of departure for developing the categories and classifications structures. Findings from a small pilot study of network users is presented to strengthen the claims of the paper.

The incredible growth of the Internet and its resources in the last five years has brought the existence of networked resources to the awareness of millions of new users. We are in a very early stage in network development, and we need a perspective of the tasks ahead. Manuscript and book publishing have been with us for centuries; so have the efforts at cataloging, classifying, and organizing those materials. Bibliographic control techniques have evolved to deal with printed materials and the information they contain. Similarly, the organizational problems of networked resources will not be resolved immediately. In addition, networked resources will mutate and emerge the network, its users, and its uses evolve. Therefore, while we must develop tools that answer present needs, these tools must evolve to accommodate an unknown future universe of networked resources.

Terminology

For this paper, network refers to any computer network, from local area networks (LAN) to national high speed data networks such as the Internet or the emerging National Research and Education Network (NREN). Networked resources can be available at any of these networks levels. Lynch and Preston (1990) provide a comprehensive overview of network development related to information resources.

The term "networked resources and services" refers to entities such as supercomputers, databases, servers, and people, which are accessible via a computer network. A specific type of networked resources can be identified as "networked information resources" which include online public access catalogs (OPACS), discussion lists, bulletin boards, multi-function information services, etc. Networked information resources are the focus of this paper. Based in part on Buckland's ideas about "information-as-thing," two aspects of networked information resources are detailed in the paper (Buckland, 1991). Resources have features as "information containers" as well as "information content."

"Navigational tools" comprise resources and entities that identify, describe, and provide access information to the networked resources. Ultimately these navigational tools and resources will be networked resources themselves, living and evolving on the network. Specifically this paper assumes that there is a need to develop a navigational tool that is a networked resource. One approximation of a navigational tool is the online public access catalog in a library.

Background

Some navigational tools for networked resources are being produced by various groups and individuals. The guides and locators are available in different forms, e.g., electronic files, printed lists, and databases (Ryan, 1991). Among the familiar guides to networked resources are:

The Internet Resources Guide (National Science Foundation, 1991)
Internet-Accessible Library Catalogs & Databases (St. George & Larsen, 1990)
UNT's Accessing On-line Bibliographic Databases (Barron, 1991)
Directory of Electronic Journals, Newsletters and Academic Discussion Lists (Strangelove & Kovacs, 1991)
Zen and the Art of the Internet (Kehoe, 1992)

Examples of databases that serve as locators of networked resources are:

HYTELNET, a facility providing information about OPACS (Scott, 1991)
ARCHIE, a facility providing information about FTP archives (Emtage & Heelan, 1990).

The Coalition for Networked Information has established a Directories and Resource Information Services Working Group to deal with the issues of directory services. The Working Group brings together librarians, networkers, and vendors in a joint effort to develop a coherent strategy for providing directory services to networked resources.

Librarians have expressed interest in expanding the USMARC format to accommodate information about online and networked resources. A discussion paper developed in 1991 by the Network Development and MARC Standards Office at the Library of Congress provided an outline of data elements necessary to code descriptive and location information for networked resources (Library of Congress, 1991).

In the networking environment, the X.500 Directory Service international standard is being explored for use as a directory tool for networked resources. A Internet Draft titled "Schema for Information Resource Description in X.500" was developed in May 1991 by staff at Merit (Weider, 1991). The paper suggested a way of holding information resource description information and incorporated the data elements from the USMARC discussion paper.

The Wide-Area Information Server (WAIS) developed by Thinking Machines is another important step forward in networked resources utilization. WAIS offers a directory of WAIS databases/servers that can be queried for information about specific servers (Stein, 1991).

The foregoing is not meant to be a comprehensive list of activities but rather shows work progressing on a number of fronts to develop navigational tools. Concurrent efforts should not become isolated and communication between these developers is important if we are to benefit from their efforts.

Current guides help users to explore the wide and growing varieties of networked resources. Exploration of networked resources is interesting and valuable for its own sake. That activity provides people with the experience of navigating the network environment and also allows them to encounter the breadth of resources available. Through such explorations, new ideas for information resources may emerge and promote further evolution of the network environment. However, exploration and effective use of networked resources are not identical. Users need more than an inventory of what's out there. They need functional tools to get to the information contained on the network. Organizing the networked resources (not physically but intellectually) to assist people in locating and accessing the information is the next important step. While we are at an early stage in the evolution and development of the network environment, the large and growing mass of resources could be used more effectively by larger number of people if the appropriate navigational tools existed.

The Scope of the Problem

In traditional library and information center environments, users have learned about the tools and finding aids to use when trying to locate information for research, educational, and recreational purposes. The primary tool to locate materials in a library has been the card catalog and its new manifestation, the online public access catalog (OPAC). Catalogs were developed through the efforts of librarians and others who attempted to identify, describe, organize and provide access to materials in their collections. To catalog items for their collections, library and information center staff acquire the information container and the cataloger, with the container in hand, describes, examines, and evaluates the container and its contents.

Auxiliary tools such as indexes and bibliographies provide additional pointers to information not adequately described by the library catalog. Access tools were developed over the years to deal with particular types of "information containers," namely books, serials, sound recordings, manuscripts, prints and photographs. These information containers are characterized by a common feature of relative immutability. The "information content" is embedded in the container. The description of the item and subject classification of the content could remain the same since the container and its information content did not change. Changes in information content was embodied in a new information container e.g., a new edition of a book.

Networked resources present many similar organizational and control problems for librarians and information professionals. The goals are the same, namely to identify, describe, and intellectually organize the networked resources to provide users the means to find and use these resources.

Some similarities between traditional library materials and the networked resources are obvious. Networked information content is made available in "containers." Much of the networked information is text-based in the form of reports, documents, messages and database records. Some networked resources already are segregated by topic or subject. For instance there are discussion groups on particular topics, electronic journals for specific academic disciplines, and database servers that provide specific types of information. Therefore, it's possible to begin providing subject access based on the self-organizing aspects of some of the resources. Networked resources, like library materials, have one or more locations, therefore location information must be provided to enable the user to access the information resource.

Major differences between traditional library materials and networked resources exist. Networked resources currently provide more than text-based documents. Supercomputers, interactive databases, bulletin boards, discussion lists and printers can all be considered resources on the network. Multi-media resources combining text, audio, and video will be available for network users. Still other resources will emerge that have yet to be visualized.

There are multi-function resources available on the network. Take the examples of the Colorado Alliance of Research Libraries (CARL) or MELVYL, the online catalog of the University of California. These resources provide access to OPACS and other locally mounted databases, serve as gateways to other computers, and, in the case of CARL, offer document delivery. To describe and intellectually organize these resources will require tools that provide multi-faceted descriptions and classifications.

The network environment is volatile and dynamic. While the information container may remain relatively stable, this may not always be the case, and the information content can change frequently. Robust navigational tools are necessary in this dynamic environment. Descriptions must accommodate regular and frequent updating appropriate to the resources.

The decentralized nature of the network environment must be taken into account when developing navigational tools. Traditional library material are produced in a relatively decentralized environment. Thousands of authors, editors, reporters, publishers and distributors create and disseminate materials. Libraries are a centralized point through which these materials pass in order to be made available to users. They have served as a focal point for describing and organizing these materials. Networked resources may require decentralized description and organization, possibly by the people who create these resources. To enable effective use of the networked resources, however, a logically centralized (but physically distributed) navigational resource may be the appropriate model. The X.500 Directory Serice standard provides the resource for such a resource (Planka, 1990). Lacking such a central resource would be similar to lacking libraries and their catalogs. Users would have to contact authors and publishers directly to see if they provided materials to satisfy the users' information request.

A computer network is complex, consisting of a variety of computer hardware, software, and telecommunications facilities. When one considers the Internet environment with its interconnection of

multiple and heterogeneous computer networks, the environment becomes more complicated. Different protocol suites of the connected networks support different applications. Operating systems and commands vary by platform. Demands are placed on users just to move around on the network. The issues and problems related to this aspect of the network environment will not be dealt with in this paper. Yet one question must be raised. What levels of network skills should be assumed of users as we develop tools to help them find information on the network?

Four Domains of the Network Environment

An essential element in developing category and classification structures is delineating the domain to be covered. The network environment appears to include the following four domains: the domain of the applications; the domain of the information containers; the domain of the information content; and the domain of the user. Navigational tools should address these four separate but interacting domains if they are to provide users with effective access to networked resources. These domains are dynamically related and integral to network use.

Domain of the Applications This domain concerns the applications supported by the network for accessing and using networked resources. There are currently three general applications: remote login, file transfer, and electronic mail. In the Internet environment for networks running TCP/IP protocols, these applications are called Telnet (remote login), FTP or file transfer protocol (file transfer), and Electronic Mail. Open Systems Interconnection (OSI) supports similar applications.

Domain of the Information Container This domain can be described on the basis of the system that is providing the resources. This is the level of the "container." There is a close relationship between these containers and the applications used to access them. For example, discussion lists are primarily based on electronic mail. Often these are called listservs, short for listservers and named for the Listserv software used. The "container" is a software program that accepts messages and redistributes them to a list of subscribers as electronic mail messages. In the case of online public access catalogs, the Telnet application is used to remotely connect to the catalog, and once connected, the container is searched in an interactive manner. Another variety of containers are those systems which hold files of information that one can pull across the network to a local system. An example of this is the Network Information Centers. FTP enables one to connect to the remote system, look at the listing for the various files, and then select one or more files to transfer to the user's local machine. There is a variety of containers housing information and these include bulletin boards, data archives, informational servers, etc. Systematically identifying and describing containers can be a first step in intellectually organizing them. Providing useful categories offer users a way of assessing what type of information resource is suitable for their information needs. Arriving at definitions, descriptions, and categories may be difficult, especially since the universe of resources is not completely known or stable. This may be the most problematic domain to address.

The information container refers to a variety of entities. They are not themselves information but hold the information content in which the user is interested. They can be referred to as information-as-thing, potentially informative objects (Buckland, 1991, pp. 42-54). For an experienced network user, knowledge of the types of containers and their relation to the services that enable access provides the basis for effective use of the resources. One learns over time what type of networked resource is best suited to particular information needs. Patrons of libraries learn where to look for information. They understand that daily newspapers, weekly news magazines, annual almanacs all contain information. Yet they each serve a distinct purpose in providing specific types of information. Similarly, patrons in libraries understand that different formats of materials will contain different kinds of information. A collection of portrait photographs will serve a different function than a monograph on genealogy, yet they both may be very important in filling out information gaps when one is trying to write about and describe members in a family history. Understanding the formats of library materials and their special features give users power in searching for

information. The universe of information containers in the network environment includes: discussion lists, newsgroups, archives of discussion lists and newsgroups, bulletin boards, multi-function services (e.g., CARL, MELVYL), directory services, informational servers (e.g., Weather Underground), data archives.

Why is it important to distinguish these information containers? Each has characteristics and features the knowledge of which can assist a user to assess its suitability for a particular information need. For example, a user wants to know about current weather conditions. The network might contain a number of resources that deal with weather information. WEATHER-L (hypothetical) is a discussion list on climate changes; WEATHER-ARC (hypothetical) is an archive of meteorological datasets; and WEATHER UNDERGROUND (actual) is a database of current weather conditions. The choice to use a particular resource will be based on the type of information the user wants. Knowing the types of information containers helps the user to narrow the choice to the most appropriate container.

The Domain of Information Content This domain refers to the actual information or data in an information container. It deals with the "aboutness" of the content. In a library or information center, users have two primary paths to get to information. One is the known item request where users look for items based on knowledge of a particular author's name, a specific book or journal title. In the network environment, a user should be able to search a navigational tool for "WEATHER UNDERGROUND" and receive access and other information to that specific resource. A question can be raised whether a known item search is a search for a container or for content. The line is somewhat fuzzy.

The other approach to information is a subject or topic search. A user may not know a specific item, but wants information about a topic such as "weather conditions." Library systems have several ways to pursue such an information query. Navigational tools for networked resources will also need to accommodate such queries. The navigational tools should accommodate a user entering a subject search for "weather conditions." One way of organizing networked resources for this type of query is to provide broad subject descriptors for information content. Investigating classification theory and library classification practice can inform this aspect of developing navigational tools.

The Domain of the User Users of networks come in increasingly varied shapes and sizes. Early users of electronic networks were small in number and relatively homogeneous -- computer specialists, technologists, scientists and researchers. As more and more networks are connected to a national computer network, the user population will become more diverse with different interests, skills, knowledge and information needs. Navigational tools for the new generation of users must accommodate these differences. However, it may be that no one navigational tool can accommodate the complete range of users. However, without addressing the real needs of users and their information behaviors, development of navigational tools will short-sighted at best, and ineffective at worst.

Libraries can be examined for both their strengths and weaknesses in the context of the users. Library catalogs reflect an intellectual organization of knowledge and description of materials to help users in locating information. Unfortunately these catalogs, based on rules for bibliographic description and classification, do not always enable a user in finding pertinent information. In part this stems from constraints of resources and inadequate ways of representing information to help user locate what is needed. It may also be due to a lack of understanding user information behaviors. A user often has to interpret or translate an information request into the structure of the classification system of a library. Online public access catalogs provide new ways of gaining access to library materials. Many systems provide keyword searching of OPAC records; a user can enter uncontrolled vocabulary terms to gain access to materials. Keyword searching will not uncover all relevant materials, but likewise the use of a particular Library of Congress Subject Heading may not uncover all relevant materials for the user either.

As developers of navigational tools for networked resources, we must focus on the users and their information behaviors in arriving at suitable tools. Taylor advances a set of eight classes of information use,

"generated by the need perceived by users in particular situations." (Taylor, 1990). These classes can be helpful when thinking about the navigational tools from the perspective of users' information needs. The classes are not meant as mutually exclusive and in fact may be inter-related.

Enlightenment: the desire for context and information or ideas in order to make sense of a situation;
Problem Understanding: more specific than enlightenment; better comprehension of particular problems;
Instrumental: finding out what to do and how to do something;
Factual: the need for and consequent provision of precise data;
Confirmational: the need to verify a piece of information;
Projective: future oriented, but not related to political or personal situation;
Motivational: has to do with personal involvement, of going on (or not going on);
Personal or Political: has to do with relationships, status, reputation, personal fulfillment.
(Taylor, 1990)

Knowledge of research findings and literature concerning users' information behaviors is required for developing adequate navigational tools.

Users and the Organization of Networked Resources

To gain an understanding of how users currently deal with the lack of navigational tools for networked resources, a small pilot study was conducted in Fall, 1991. Interviews with eight network users discussed how they learn about networked resources, what resources they use, and what they consider helpful or necessary in navigational tools.

Most of the users felt that the network environment itself demands a lot of effort just to use. Several users mentioned the need for better interfaces, common or shared commands across platforms, and a generally less complicated environment. This is important since developing navigational tools is only one aspect of navigating the network. Can navigational tools be developed that minimize some of the other difficulties of getting around the network?

These users have found ways for gathering information about networked resources. Knowledge of the existence and location of resources often is passed among other users. Messages circulated on discussion lists are another source for information about resources. However, these users rely mostly on known resources and familiar functions in their use of the network. A person who uses newsgroups to gather specific information tends to operate in that environment for a range of information activities.

The users have developed an understanding that different networked resources serve different information needs. For instance, if a user has a non-urgent need for information about a topic, he or she might subscribe to a discussion list on the topic and passively receive messages. If however, information is needed quickly, the user might post a query on that discussion list, actively seeking information or pointers to information for an answer. Users noted that it helped to know whether the discussion list had an archive of messages that could be examined for pertinent messages related to the information request. For this type of networked resource (a discussion list), several user behaviors were exhibited and different aspects of the resource were important from the user's perspective. One first must find an appropriate discussion list for the topic; the urgency and nature of the information request suggest different patterns of actions; and knowing whether there is an archive of older messages affects selection of a discussion list.

The fact that respondents indicated that facets or attributes of networked resources are important (e.g., knowing whether or not a discussion list is archived) for evaluating its usefulness confirms the need

to delineate these resources' features and characteristics. Seeing the similarities and differences among resources is part of a categorization process.

Users stated they would like to make a query of a navigational tool and be pointed to appropriate networked resources. They want to submit a subject, keyword, or term search and receive back pointers to networked resources that contain or deal with that subject or topic area. They suggested that knowing whether the resource was an FTP site, a discussion list or newsgroup (and whether it had archives), or a telnet resource was important in determining which they would use. This support this paper's premise that we must be able to identify, describe and categorize the networked resources at the information container level. In addition, there must be a mechanism to allow subject/keyword searching of a navigational tools to arrive at those networked resources that are pertinent to a request.

Another important idea resulting from the interviews concerned the need for "mental models" and "visualizations" of the applications and resources and how they fit in the network environment. Based on their interaction with devices in the environment, people form mental models of the device "largely by interpreting its perceived actions and its visible structure" (Norman, 1988, p. 17). When the functions of the application and characteristics and features of networked resources are better understood i.e., users have a mental model of what is happening, people will can more effectively use those resources and applications. One respondent mentioned the telephone as an example. We have an understanding (mental model) of how a telephone "works." This understanding is not at the technical level, nor does it need to be. We know how to "use" the telephone based on a non-technical notion of how it "works." Similarly we have models of library materials that help us "use" them. The mental model of a book or magazine helps us choose the one which best serves our information need. Mental models for the network environment are necessary if users are to be successful in searching the varieties of resources of the network.

The exploratory mode in which the respondents operate suggested they were more interested in finding what's "on the net" rather than looking for specific information. When they look for specific information now, they are likely to use what they know about existing resources and base their query on that knowledge e.g., contacting a discussion list that deals with a specific topic and asking for information. In some cases they have discovered or heard about a resource that is sufficient for current information needs e.g., a file transfer resource with programs for a particular kind of computer.

The respondents were knowledgeable about the applications and some of the resources, but were uncertain about categorizing them. They know what each of the applications (electronic mail, remote login, file transfer) can do and that the applications are fundamental in navigating the network. Details about specific resources, such as an archived discussion list, is desirable, but the more important level of categorization and classification for the respondents was at the subject or topic level.

Classification Issues in the Domain of Information Containers

Categorizing, developing taxonomies, and creating classification schemes reflect a human propensity for meaningfully organizing items in our experience. These processes help us navigate through the world and communicate with others about our experiences and things in the world.

> Categorization performs a fundamental function in the process of cognition. By recognizing the similarities between potentially dissimilar entities, the individual is enabled to form theories, or models, of her environment that allow her to extend to new encounters the generalizations garnered from past experience (Jacob, 1991, p. 75).

Categorizing the information containers of networked resources is a first step in developing navigational tools. This process may also help provide a basis for developing the mental models referred to above.

Librarians and information professionals have used a variety of techniques, procedures, and vocabulary to organize information resources. Materials have been organized by format, such as audio-visual, monographs, serials, sound recordings. Content of materials has been intellectually organized through sophisticated classification structures such as Library of Congress Classification, Dewey Decimal Classification, and Universal Decimal Classification systems. These organizational techniques assume entities have certain properties and attributes (physical or intellectual) by which they may be described and grouped together. Placing items in categories is done by seeing similarities or patterns among items and grouping them based on perceived characteristics. A subsequent step can be the development of a classification structure in which the categories are related to one another. One category may be related to another category because it exhibits a "kind of" or "part of" relationship to other items. Classification schemes pull together the categories into relationships and designate those relationships. A taxonomy can describe a fully developed classification scheme and show the relationships among the classes. The classification structure is a potentially informative object providing information about what is classified and the relations among the entities.

Returning to the domain of information containers, how might categories of these entities be set up? Given the complexities of what comprises an information container, a further delineation of this domain may be helpful. The domain can be separated into "system information containers" and "document information containers." System information containers can include: discussion lists, newsgroups, archives of discussion lists and newsgroups, bulletin boards, multi-function services (e.g., CARL, MELVYL), directory services, informational servers (e.g., Weather Underground), and data archives. These "system information containers" will hold one or more of the "document information containers." A partial list of "document information containers" includes: reports, abstracts, indexes, articles, guides, minutes, etc.

Useful categories for information containers should not be considered mutually exclusive. It might be more effective to provide facets that can be associated with the categories. Facet analysis allows attributes and features of the containers to be added or removed as the container changes. This will be a helpful and flexible approach as new varieties of information containers evolve on the network.

It has been asserted that unlike traditional library materials, networked resources are dynamic and volatile. Generally this means the information content changes rather than the container itself. However, information containers themselves can take on new features and attributes. Take for example a discussion list such as USMARC-L, a listserv-based discussion list set up by the Library of Congress Network Development and MARC Standards Office in conjunction with the University of Maine. The list is used by people to discuss USMARC format development, maintenance, and implementation. USMARC-L includes an archive of messages passed among subscribers. Recently files of discussion papers, minutes of meetings, agendas and other information related to USMARC were made available. These documents are not sent out as electronic messages by the listserver but announced in messages from the list moderator. Subscribers are directed to use the facilities of the listserv software to retrieve the files containing these documents. This is an example where a new feature (file transfer) was implemented in the system information container and new document information containers were added to it.

A common vocabulary of attributes, features, and facets is needed to describe information containers. The list of terms must be open-ended to accommodate new containers, new features, and new attributes. Terms that resonate for network users based on their experience may facilitate the development of mental models. While "electronic mail" may be a helpful term, "listserv" may not.

Should a complete classification scheme be developed for information containers? Possibly. The reason to take the analysis one step further into classifying, as opposed to inventorying or categorizing is that some of these containers may be related to other containers. For example, an archived discussion list is related to a discussion list in a "kind of" relationship. The classification structure becomes an informative object and may provide guidance in selecting and using the resource.

Moen

Summary

This paper has identified a number of issues and problems that need to be addressed in the development of navigational tools for networked resources. Parallels and differences between traditional library materials and networked resources were presented. Bibliographic organization techniques that have helped librarians and information professional with organizing their collections can inform the organization of networked resources. Categorization and classification techniques must attend to the quite different nature and characteristics of the network environment and its resources. It was proposed that there are four domains (applications, information containers, information content, and users) of the networked environment to take into account when developing navigational tools. Users' behaviors and needs are an important source for information to guide the choice of navigational tools.

The network environment is relatively new, has not matured or stabilized, and will evolve new networked resources. Thus, any guidelines for navigational tools must emphasize an open-endedness to the tools. They must be expandable and flexible. One limitation of classification schemes used in library practice has been the restrictions in accommodating new areas of knowledge and in some cases, new formats for information. While we are able to develop new navigational tools informed by traditional activities of organizing and classifying information, we should be aware of their limitations and constraints. This is very important given the rapidly expanding universe of networked resources. Robust navigational tools that fit the dynamic network environment and its users are fundamental to the effective use of networked resources.

References

Barron, Billy <billy@untvaxb.acs.unt.edu> (1991). *UNT's accessing online bibliographic databases*. Denton, TX: University of North Texas. Available: FTP vaxb.acs.unt.edu <129.120.1.4> Directory: LIBRARY File: LIBRARIES.TXT

Buckland, Michael K. (1991). *Information and information systems.* New York: Greenwood Press.

Emtage, Alan <bajan@cd.mcgill.ca> & Heelan, Bill <wheeler@cs.mcgill.ca>. (1990). *ARCHIE: The McGill school of computer science archive listing service*. Montreal, Quebec: McGill University. School of Computer Science.

Jacob, Elin K. (1991). Classification and categorization: drawing the line. *Proceedings of the 2nd ASIS SIG/CR Workshop on Classification Research.* (pp 63-79). Held at the 54th ASIS Annual Meeting, Washington, DC. October 27, 1991.

Kehoe, Brendan P. <brendan@cs.widener.edu>. (1992). *Zen and the art of the Internet: a beginner's guide to the internet.* Chester, PA: Widener University. Available: FTP ftp.cs.widener.edu <147.31.254.132> Directory: PUB/ZEN Files: ZEN-1.0.tar.Z, ZEN-1.0.DVI, and ZEN-1.0.PS

Library of Congress. Network Development and MARC Standards Office. (1991). *Dictionary of data elements for online information resources* (Discussion Paper #49). Washington, DC: Library of Congress.

Lynch, Clifford A. & Preston, Cecilia M. (1990). Internet access to information resources. In Martha E. Williams (Ed.) *Annual Review of Information Science and Technology*, (pp. 263-312). Vol. 25. Medford, NJ: Learned Information.

National Science Foundation. Network Service Center. (1991). *Internet resource guide.* Cambridge, MA: NSF Network Service Center. Available: FTP NNSC.NSF.NET <192.31.103.6> Directory: RESOURCE-GUIDE

Norman, Donald A. (1988). *The psychology of everyday things.* New York: Basic Books.

Planka, Daniela. (1990). Network directory services. *Library Hi Tech* 8(4). 93-103.

Ryan, Joe (1991). Resource review. *Electronic Networking* 1(1). 40-46.

St. George, Art <stgeorge@unmb.bitnet> & Larsen, Ron <rlarsen@umd5.umd.edu>. (1990). *Internet-accessible library catalogs and databases.* Albuquerque, NM: University of New Mexico. Available: E-mail: LISTSERV@UNMVM.BITNET Message: GET LIBRARY PACKAGE

Scott, Peter <scott@sklib.usask.ca>. (1991). *Hytelnet.* Saskatoon: University of Saskatchewan. Available: FTP vaxb.acs.unt.edu <129.120.1.4> Directory: LIBRARY File: HYTELN30.ZIP

Stein, Richard Marlon. (1991). Browsing through terrabytes. *Byte* May 1991. 157-164.

Strangelove, Michael & Kovacs, Diane. (Compilers). Okerson, Ann (Ed.). (1991). *Directory of electronic journals, newsletters and academic discussion lists.* Washington, DC: Association of Research Libraries.

Taylor, Robert S. (1990). Information use environments. *Progress in Communication Science* 10. Norwood, NJ: Ablex Publishing Co.

Weider, Chris. (1991). *Schema for information resource description in X.500*. Internet Draft available from NSF Network Service Center.

Access to Electronic Information Resources within USMARC

Rebecca Guenther
Senior MARC Standards Specialist
Library of Congress

This paper discusses efforts to accommodate electronic information resources within the USMARC formats. The Library of Congress has been exploring this issue by establishing a framework for discussion. In addition, it has compiled a list of data elements needed to accommodate online information resources.

Standardization is necessary for several reasons: to enable systems to provide more direct access to networked information resources using machine processing of data; to identify a set of required data elements that give sufficient information about the resource for efficient access to it; and to share data between different types of systems.

The paper discusses the specific data elements needed for description and access to online information resources. It divides this group of information into two parts: electronic data resources and online systems and services. It reviews the data elements available in the USMARC computer files format and discusses how it accommodates the two groups.

Electronic data resources consist of computer software, documents stored in machine-readable form, databases of bibliographic or numeric data, directories, etc. They may exist in different formats and may be accessible via multiple online systems or FTP sites. Data elements required for description and access to electronic data resources and their accommodation in USMARC is reviewed.

In addition to identifying new types of data elements required for electronic data resources, the definitions and scope of those that exist in USMARC for computer files are reconsidered in terms of how this type of information differs from traditional bibliographic items. In particular, the issue of how to determine the location of the resource and subsequent access to it is discussed. The problem of how to describe multiple forms of the electronic data resource is considered.

Online systems and services include library information systems, commercially available systems (e.g., DIALOG), community- wide information networks (e.g., Freenet), etc. For this group, the concepts of describing traditional bibliographic data do not necessarily apply, and the computer files format is not adequate, particularly for access information. Data elements required for online systems and services in USMARC is reviewed. Those that are available in the new provisionally approved USMARC community information format are noted.

Problems in describing and providing access to online systems and services are reviewed. How a record for an online system could provide enough information and in what form to allow for an automatic login into that system is considered.

How electronic data resources and online systems/services might interrelate in an online system is reviewed with a model suggested. There are some types of online information resources, such as bulletin boards and Listservs, that do not easily fall into one or the other category. How to integrate the two groups, particularly in terms of these that do not fall into one or the other, is considered.

Providing access to electronic information resources within USMARC changes the traditional notion of the library catalog. How to maintain this dynamic information that is constantly changing may prove to be a difficult challenge.

Multimedia as Rhizome:
Design Issues in a Network Environment

Kathleen Burnett
Department of Library and Information Studies
School of Communication, Information and Library Studies
Rutgers University
New Brunswick, NJ

Kathleen Burnett
School of Communication, Information & Library Studies
Rutgers University

Multimedia as Rhizome: Design Issues in a Network Environment

1.0 Introduction

> While the study of the temporal and spatial distanciation of communication is important to the concept of the mode of information the heart of the matter lies elsewhere. For the issue of communicational *efficiency* ... does not raise the basic question of the *configuration* of information exchange .... (Poster 1990:8)

The purposes of this paper are twofold: 1) to establish a working vocabulary comprised of a set of well-defined terms which will enable an intelligent discussion of multimedia network design; and 2) to lay one of many possible foundations for that discussion through an exploration of a theory of hypermedia design, particularly as it might relate to a design of multimedia networks.

What distinguishes hypermedia design from that of other modes of information is not that it is computer-driven--afterall, the computer played no role in Vanevar Bush's memex vision--nor that it is interactive, since the entire history of oral communication, whether electronically mediated or not might be characterized as interactive; nor even that it includes navigational apparatus such as "links" and "nodes," which might better be thought of as "symptoms" than "causes," or "buttresses" rather than "groundwork"; but that it posits an information structure so dissimilar to any prior human communication system that it is difficult to describe as a structure at all. It is not linear, and therefore may seem alien when compared to the historical path written communication has traversed; it is not hierarchical nor "rooted," and therefore may appear chaotic and entropic. As information transfer structures, our current computer networks share many similar characteristics. From a bird's-eye perspective they are alinear, non-hierarchical, and bulbous. In their architecture and patterns of growth, networks have more in common with botanical forms than the information transfer structures of the past.

1.1 Definitions

For both historical and theoretical reasons, the literature of definition in the multimedia area is full of ambiguity and prevarication. Terms such as "multimedia," "hypermedia," and "hypertext" are used both interchangeably and to mean very different things. Today, one vendor's "multimedia application" is another's clip-art. Standards are still in development. Applications appear and disappear so rapidly that it is difficult for even the informed user to keep up to date. Even if we follow the lead of some enthusiasts and date the inception of the field to the publication of Vanevar Bush's memex vision, we are only talking of a scant history of fifty years, including a period of dormancy of at least twenty years. In terms of active development and discussion we can probably claim no more than fifteen years, peaking within the last five.

The term "networking"is ambiguous as well, depending on the context of its use, which are as various as social interaction (e.g., a network of individuals sharing information), transportation (e.g., a system of railways, roads, canals, etc.), and electronics (e.g., a system of connected electrical connectors). According to the *Concise Oxford Dictionary*, 8th ed., a "computer network" may be defined as "a chain of interconnected computers, machines, or operations;" and "computer networking," as "link[ing] (machines, esp. computers) to operate interactively." My interest in computer networking is perhaps at once more limited and yet broader in its implications than the scope of this definition. I am really only interested in the interactive information transfer function of computer networks, and yet this seeming limitation points out the narrowness of the definition itself. Computer networks interest me less as structures in and of themselves, than they do as structures which enable human-computer and computer-mediated human-human interaction, a dimension which , it seems to me, is completely absent in the definition .

To a certain extent, this definitional confusion may be explained by the relative youth of both multimedia and computer networking. In academia today we rely so heavily on computer networks that we forget that they didn't exist thirty years ago. Ten years from now multimedia networks will be so ubiquitous that we will have forgotten the days when most networks could not handle images, sounds and dynamic media.

The second reason for the confusion is probably the more interesting because it is the more systemic. Whatever you call these new modes of information and information transfer, they are, at their cores, amalgams. "Multimedia" allows the combining, co-mingling and intertwining of media we are accustomed to thinking of as separate, while simultaneously supporting synchronous and asynchronous, hierarchical, historical and synthetic structures. It is, at its very roots, a multiple, and therefore may appear amorphous and even chaotic. A visual survey of contemporary maps and representations of network architectures reveals similar characteristics and patterns.

1.1.1 Hypertext

The organizing principle for these multiples is usually referred to as "hypertext," an amalgam in itself made up of the prefix "hyper-," derived from the Greek "φπερ," meaning over or beyond, and of the common English word "text." According to the *Concise Oxford Dictionary*, 8th ed., "text" is defined as :

> 1. the main body of a book as distinct from notes, appendices, pictures, etc.; 2. the original words of an author or document, esp. as distinct from a paraphrase of or commentary on them; 3. [Relig]a passage quoted from Scripture, esp. as the subject of a sermon; 4. a subject or theme; 5. (in "pl."[Education]) books prescribed for study; 6. [US][Education] a textbook; 7. (in full "text-hand") a fine large kind of handwriting esp. for manuscripts.

The third through seventh definitions all deal with specific contexts; the first and second are too restrictive to provide a foundation for understanding what Ted Nelson had in mind when he coined the term in the sixties. An examination of the roots of the term "text," however, reveals an interesting interplay. "Text" derives ultimately from the Latin *texere* which had nothing to do with books or writing, but rather with weaving, hence the English "textile." Recent publications (Vandergrift, etc.) have discussed the net-like structures developed using HyperCard®, a popular hypermedia application. I like the sense that this lends to the meaning of "hypertext" as an art "beyond weaving," allowing for infinite variation in color, pattern, material and structure. It is unfortunate that this is not the way the term is commonly understood, because it gets to the heart of what it signifies. I would propose, then, that "hypertext" be understood as the organizing principle

of "hypermedia," rather than being used to describe applications or groups of applications which make use of navigational tools such as links and nodes to form textual databases. Nor should it be used as an appellation for a database which uses such devices. Along the same lines, while a multimedia network should not be referred to as a "hypertext network," "hypertext" should be present as a meta organizational principle in any true network of this type.

1.1.2 Hypermedia

The term hypermedia was borne out of a misunderstanding of the meaning of "hypertext." Hypermedia is generally used in two ways: 1) to describe applications which make use of navigational tools such as links and nodes to form mixed media databases, and 2) to describe the organizational principles of such databases. As far as I can discern, "hypertext" as an organizing principle lacks nothing that might be required, or even desirable, in the production of databases in non-textual media, therefore I find the second usage redundant and confusing. On the other hand, I see no reason that "hypermedia" should be limited to the description of mixed media databases--why not include single media databases which partake of the organizational principle of hypertext? I would propose, therefore that use of the term "hypermedia" be limited to the description of databases in any media which have hypertext as their organizing principle.

1.1.3 Multimedia

"Multimedia" is perhaps the least precisely used term of the three. At times it is used as a synonym for "hypermedia," at others as a kind of antonym, implying analog production in mixed media. I would like to propose a definition which allows for both meanings. Multimedia, as I understand it, is a generic term encompassing the use of multiple media, digital and analog, in a variety contexts. Therefore, some "hypermedia" is "multimedia"-- any database that stores information in more than one media and employs "hypertext" as an organizing principle is "multimedia," but a single-media "hypermedia" database is not. On the other hand, relational and flatfile databases which store information in more than one media are "multimedia" but not "hypermedia" since they do not use "hypertext" as an organizing principle (cf. fig. 1).

1.1.4 Multimedia Networks

Multimedia networks differ from standard computer networks in that they allow for the transference of both digital single media and digital multimedia. They must facilitate the transference of flatfile and relational databases as well as hypermedia and single-media hypertext databases. This should not be problematic since hypertext may serve as a meta organizational principle. Current hypermedia applications allow for the nesting of flatfile and relational databases within hypermedia databases, making use of hypertext as a meta organizational principle. In multimedia networks, a similar kind structuring will be necessary, utilizing hypertext as a meta organizational principle for the transference of digital signals of information in any medium or combination of media no matter how the information is organized (cf. fig. 2).

2.1 Theory of hypermedia design

> By mode of information I similarly suggest that history may be periodized by variations in the structure in this case of symbolic exchange, but also that the current culture gives a certain fetishistic importance to `information.'
>
> Every age employs forms of symbolic exchange which contain internal and external structures, means and relations of signification. Stages in the mode of information may be tentatively designated as follows: face-to-face, orally mediated exchange; written exchanges mediated by print; and electronically mediated exchange (Poster 1990: 6).

Any theory of hypermedia design must support the coterminous existence of visual, verbal and combinatory modes of information. While these modes may exist in current databases, let me emphasize that what follows is a discussion of theory, not a representation of characteristics found in currently available applications calling themselves "hypermedia."

In *A Thousand Plateaus*, Deleuze and Guattari offer the following description of their third type of "book," the type which appears to be the rough equivalent of Poster's fourth stage in the mode of information, "electronically mediated exchange:

> A system of this kind could be called a rhizome. A rhizome as a subterranean stem is absolutely different from roots and radicles. Bulbs and tubers are rhizomes. Plants with roots or radicles may be rhizomorphic in other respects altogether .... Burrows are too, in all their functions of shelter, supply, movement, evasion, and breakout. The rhizome itself assumes very diverse forms, from ramified surface extension in all directions to concretion into bulbs

> and tubers .... The rhizome includes the best and the worst: potato and couchgrass, or the weed (6-7).

Telecommunications systems are rhizomorphic, as are computer networks. Think of maps you have seen and descriptions you have heard of the Internet--a rhizome. If we accept the rhizome as a metaphor for "electronically mediated exchange," then hypermedia is its apparent fulfillment, and Deleuze and Guattari's "approximate characteristics of the rhizome"--principles of connection, heterogeneity, multiplicity, asignifying rupture, and cartography and decalcomania--may be seen as the principles of hypermedia design.

2.1.1 Principles of connection and heterogeneity

The principles of connection and heterogeneity state that "any point of a rhizome can be connected to any other, and must be (Deleuze & Guattari 7)." In this sense a rhizome is very different from a tree structure, where the order is fixed by a hierarchy of relationships. Cognitive jumps, which must be mechanically forced in an hierarchy, are intuitively sustained in a rhizome.

A rhizome is the only structure which can effectively sustain connections between different media without giving hegemony to language. Many current relational and flatfile multimedia database applications support the storage of multiple forms of media, and some will even display different types contiguously, but key*word* searching is the only mechanism provided for cross-type searching. The meaningful formation of hierarchies across media boundaries can only be accomplished through the use of language, since hierarchy is itself a creation of language, language is the only universal tool available within an hierarchical structure. A rhizomorphic structure, on the other hand, does not rely on language for its ordering, although many of the linkages in a given structure may be linguistic.

> A rhizome ceaselessly establishes connections between semiotic chains, organizations of power, and circumstances relative to the arts, sciences, and social struggles. A semiotic chain is like a tuber agglomerating very diverse acts, not only linguistic, but also perceptive, mimetic, gestural, and cognitive; there is no language in itself, nor are there any linguistic universals, only a throng of dialects, patois, slangs, and specialized languages.

Hypermedia design is rhizomorphic in its sustenance of heterogeneous connection, because there is no systemic hierarchy of connection. The perception of connectivity is entirely left to the user,

though the pre-existence of particular connections may foster varying user perceptions of overall structure. Current computer networks show similar patterns of development, for while the architecture of individual LANs and WANs may be perceived as linear or hierarchical, the connection to national and international networks disconcerts such perception by altering its context. The tendril of connection, be it direct phone line or gateway, reorganizes the perception of the local architecture by rendering it on a map of a much larger territory.

2.1.2 Principle of multiplicity

> *"All things tend to decenter ...."* (Gertrude Stein, *Tender Buttons)*
>
> A multiplicity has neither subject nor object, only determinations, magnitudes, and dimensions that cannot increase in number without the multiplicity changing in nature .... An assemblage is precisely this increase in the dimensions of a multiplicity that necessarily changes in nature as it expands its connections. There are no points or positions in a rhizome, much as those found in a structure, tree or root. There are only lines. (Deleuze & Guattari 8)

Hypermedia design is able to support non-hierarchical thinking and cognitive jumping because it recognizes the diversity of multifarious modes of information. Information may be structured hierarchically within a hypermedia system, but only to the extent that such a structure exists in a coterminous relationship with other structures. In other words, hypermedia design presupposes not only that multiple points of access are preferable to a single point, but by extension, that multiple structures are preferable to a single structure. Information retrieval studies have shown that a single user's selection of access points for a given topic may vary over time and space, making it difficult for an indexer to predict potential user vocabulary. The principle of multiplicity is reflected in hypermedia design by the coterminous presence of varying modes of access to a single structure, on the one hand, and of varying structures on the other. In the case of multimedia networks, the application of this principle is even more essential since such structures must support multiple points and forms of access not only across space and time, but also coterminously and simultaneously.

2.1.3 Principle of asignifying rupture

Hypermedia design intuitively supports two forms of access which must be forced in

hierarchical structures: user-generated access and mapping. The principle of asignifying rupture supports the former, and those of cartography and decalcomania, the latter. In an hierarchical structure, a user-generated access point may cause a rupture in the system. For example, a user may, through the process of serendipity, arrive at a particular point in a hierarchy, even though her departure-point has no apparent hierarchical relationship to that arrival point. If she is allowed to introduce the departure term into the hierarchy without further evaluation, the very structure of that hierarchy might well be undermined. In contrast, hypermedia design encourages such "disruptive" activity while rendering it insignificant. Since the structure does not rely on any given theory of relationship, it cannot be affected by the characterization of a new relationship previously alien to it. The potential for any relationship exists within hypermedia; some simply await unmasking.

2.1.4 Principles of cartography and decalcomania

The second form of access not easily supported within an hierarchy is mapping. Tracings or logs of an individual's progress through an hierarchical database are of course possible and may help a user to "retrace" a given path, or provide useful data for research in human-computer interaction. Deleuze and Guattari's notion of mapping is, however, quite different, and presupposes the operation of the principles discussed previously.

Each user's path of connection through a database is as valid as any other. New paths can be grafted on to the old, providing fresh alternatives. The map orients the user within the context of the database as a whole. In hierarchical systems, the "user map" generally shows the user's progress, but it does so out of context. A typical search history displays only the user's queries and the system's responses. It does not show the system's path through the database. It does not display rejected terms, only matches. On additional command, it may supply a list of synonyms or related terms, but this is as far as it can go in displaying the "territory" surrounding the request. It can only understand hierarchy, so it can only display hierarchical relationships.

> What distinguishes the map from the tracing is that it is entirely oriented toward an experimentation in contact with the real. The map does not reproduce an unconscious closed in upon itself; it constructs the unconscious. It fosters connections between fields, the removal of blockages on bodies without organs, the maximum opening of bodies without organs onto a plane of consistency. It is itself a part of the rhizome. The map is open and connectable in all of its dimensions; it is detachable, reversible, susceptible to constant modification (12).

A hypermedia map is more closely related to geographic maps than to search histories. It shows the path of the user through the surrounding territory. Some of that territory is "charted"--it is well mapped out in terms that the user understands, and connected to familiar territory or nodes, and some is "uncharted"-- either because it consists of unlinked nodes that exist in the database much as an undiscovered island might exist in the sea, disconnected from the lines of transfer and communication linking other land areas; or as an unidentified planet in space, with the potential for discovery and even exploration, but as yet just a glimmer in the sky--or because it is "linked" in ways that are meaningless to the user in his present context. The user can zoom in on zones of interest, jump to new territories using previously established links or by establishing new links of his own, retrace an earlier path, or create new islands or nodes and transportation routes or links to connect them to his previous path or the islands or nodes charted by others.

The principle of cartography is perhaps the most important of the principles of the rhizome to the design of multimedia networks. No structure which is bound by time and space is capable of accurately representing a network. Networks are in a state of constant growth and change; they must be infinitely flexible; they must respond to and transfer a wide range of signals instantaneously; and they must support instantaneous passage across vast distances. Traditional forms of mapping and architectural rendition can only represent the network as it is frozen in space and time. In other words, it is possible to represent the physical dimension of a network--the interactive linking of computers--and perhaps even the social dimension--the facilitation of human-computer and computer mediated human-human communication--at any given moment, but it is not possible, using traditional mapping techniques, to represent these dimensions in motion and across time.

3.0 Conclusion

Hypermedia is neither a tool nor a device; it is a way of thinking about and organizing information. Its power derives from its flexibility and variability; from its ability to transmute and transcend any traditional tool or structure. A theory of hypermedia design must be developed in order to cope with its amorphous nature, but, perhaps because of that very nature, any theory must be in and of itself amorphous. The characterization of hypermedia as a rhizomorphic structure with hypertext as its organizing principle will aid in the design of multimedia networks which share similar structural patterns requiring the implementation of hypertext as a meta organizational principle,

---

References

Deleuze, G. & Guattari, F. (1988). *A Thousand Plateaus..* London: Athlone Press.

O'Dell, Peter . (1989). *The Computer Networking Book.* Chapel Hill, NC: Ventana Press.

Poster, M. (1990). *The Mode of Information..*Chicago: University of Chicago Press.

Vandergrift, K. (1988). "Hypermedia: Breaking the Tyrrany of the Text," *School Library Journal.* 35:3 (Nov. 1988): 30-35.

Figure 1: The Domain of Multimedia

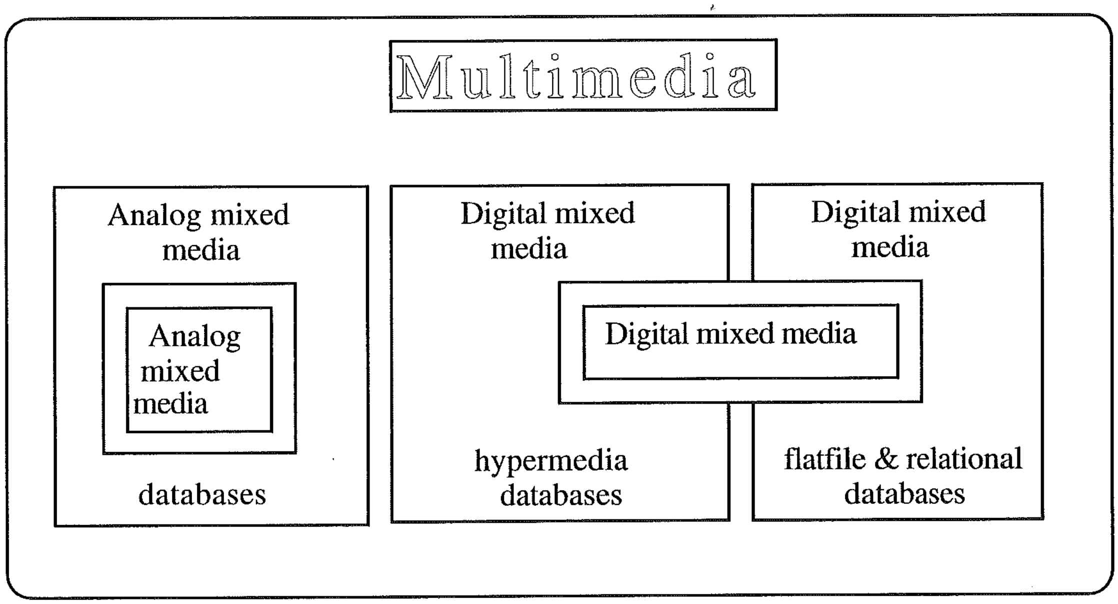
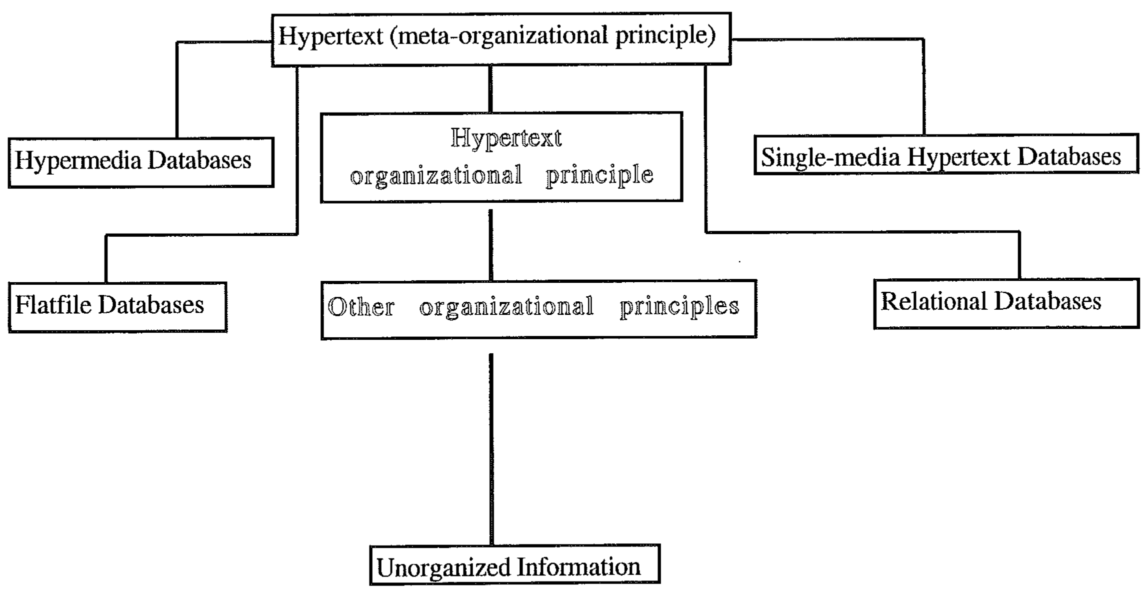


Figure 2: Hypertext as Meta-organizational Principle



DELPHI - AN INFORMATION RESOURCE IN A MULTIVENDOR MULTIPROTOCOL NETWORK ENVIRONMENT

Jeff Sedayao
Intel Corporation

Jane Starnes
Intel Corporation

Sally Hambridge
Intel Corporation

ABSTRACT

Providing a networked information resource in a multivendor, multiprotocol environment is a challenging task. Intel's network has four major network environments, each corresponding to functional areas within the company. Each of the four functional areas - design engineering, manufacturing, sales and marketing, and administration - has its own network transport and electronic mail protocol. DELPHI is a computer information resource for Intel Corporation. DELPHI provides a variety of services: a bulletin board, databases including a library catalog, internal technical memos, hazardous material handling information, and stock information. Subject-specific news is gathered, screened, posted on the bulletin board, and distributed by electronic mail. The key challenge is to make DELPHI accessible to Intel employees regardless of what environment they use. DELPHI implementors have created a "login" user interface that can be used over different transport protocols. They have also encouraged the implementation of an open and widely available protocol suite, TCP/IP, on personal computers and IBM mainframes. Since DELPHI has implemented TCP/IP, it then becomes available to users on PCs and on mainframes. DELPHI also has environment-specific interfaces to other applications such as subject specific news, automated mailing list maintenance, and stock information. DELPHI services have proved remarkably popular within Intel, so much so that DELPHI is typically overloaded during a work day. Future plans include upgrading DELPHI CPU, improving networking, and defining a long term information architecture strategy.

Introduction

The network environment of a large corporation is typically complex. Functional groups within the corporation, such as engineering, manufacturing, and marketing, have differing focuses. The different groups use hardware from different vendors. The different vendors use different network protocols. The different functional groups have different information needs. To make things even more complicated, each group's personnel may be scattered across several continents.

Such a complex and diverse network environment poses problems for an information resource provider. An application developed on one computing platform may be inaccessible by users on other platforms. Take, for example, a library catalog application on a Digital Equipment Corporation (DEC) VAX designed for remote access. The remote users must log into the VAX and use the application interactively. A personal computer's only network application might be electronic mail. With no virtual terminal/remote login protocol on such a PC, that PC's user cannot access the library catalog application. If the user moved to an IBM mainframe, he might also encounter access problems. While the DEC VAX and the IBM mainframe both may support a virtual terminal protocol, it is entirely possible (and often highly probable) that they do not support the same one. The result is the same as if the user were on the PC.

The library catalog application could be written so that the VAX, the PC, and the mainframe, share information. If a computer could not interoperate with the VAX, software would be written for it. This kind of approach forces the application designer to implement software on every kind of computing platform in the company. A company might have many different types of machines and hundreds of each type. The task of implementing, maintaining, and distributing such software is daunting.

The example above gives a brief example of the problems that can be encountered in a corporate network environment. Providing a usable and accessible networked information resource in a diverse multivendor, multiprotocol environment is definitely a challenge. To meet that challenge, Intel Corporation has implemented the DELPHI information system.

This paper describes how the DELPHI system provides information resources in the complex network environment at Intel Corporation. The first section details Intel and it's networks. It covers Intel's major functional groups, their hardware, and the network protocols they use. The second section describes DELPHI's information services. DELPHI satisfies a variety of information needs, some of which are specific to particular corporate functions. Once information is gathered, it needs to reach the people who need that information. The third section covers how DELPHI connects its information to the different people that need it. The resulting experiences are covered in the fourth section. Finally, plans for the future of DELPHI are explored in the last section.

I. The Intel Network

Intel corporation is an international manufacturer of microcomputer components, modules, and systems. Intel's corporate network is spread across Asia, Europe, and the North America, connecting more than 25,000 employees at manufacturing, research, and sales offices. Intel's network can conceptionally be divided into four functional environments: 1) Design Engineering, 2) Manufacturing, 3) Business, and 4) Administration. This section will describe each of the environments and the networking protocols and application that each one uses.

Design Engineering

The design engineering environment encompasses the design and development of new Intel products. Design sites can have hundreds of Unix workstations. These workstations are connected with networks using TCP/IP based protocols. Virtual terminal capabilities in this environment are typically provided by the *rlogin* protocol or the *telnet* protocol. The *Simple Mail Transport Protocol* (SMTP) provides a way to exchange electronic mail.

Manufacturing

Intel manufacturing environments use computers to monitor production and product quality. Manufacturing engineers primarily use DEC computers running the VMS operating system. As a result, the most important protocols at manufacturing sites are the *DECnet* protocols. The *LAT* protocol connects terminal servers to hosts. The *CTERM* protocol provides additional virtual terminal capability while the *MAIL11* protocol handles messaging.

Business

Functions in the business environment include sales and marketing, human resource, and payroll applications. These applications run on IBM mainframes, and networking is accomplished using *SNA* (Systems Network Architecure). Business users use a mainframe based electronic mail system.

Administration

Basic applications like wordprocessing and spreadsheets lie in the administration environment. These applications are done on Intel X86-architecture personal computers. The PC's are networked together using the Banyan *VINES* protocols. VINES makes file and printer sharing possible. Messaging in this environment takes place with Lotus *cc:Mail* or Banyan mail.

Without special application gateways, these four environments cannot communicate with one another. Neither electronic messaging nor remote login are possible. To complicate matters, personnel in each functional environment can be in different states, different countries, and different time zones. In this difficult environment, DELPHI was implemented to provide information services.

II. DELPHI Information Products

DELPHI information products were created to satisfy a number of information needs. Engineers, marketers, and other Intel personnel need access to technical information to do their jobs. This information exists in several forms. Some information is contained in technical forums conducted via electronic mail. Other information resides in the libraries spread across Intel. Still more information is in technical memos created by researchers and developers. This section describes the services that DELPHI provides in order to meet those needs.

One of DELPHI's first products is providing access to the technical forums available on the Internet. Developers and researchers want access to the discussion lists on such subjects as artificial intelligence, Computer Aided Development (CAD), and Computer Integrated Manufacturing (CIM). While some of the researchers and engineers choose to have the forums mailed to them directly, others want the forums put in a place that they could read at their leisure. Centralizing the lists has other advantages. Disk space across the company could be saved, and the network costs of distributing the forums around the company could be minimized. To solve this problem, DELPHI offered a bulletin board product. Messages from the Internet arrive and are subsequently posted on the bulletin board. The messages can then be read at users' convenience.

Intel has libraries in sites all over the world. For engineers and researchers to effectively use these libraries, they have to know what is available in them. DELPHI services provide on-line library and periodical catalogs. Abstracts of internal technical memos is another important DELPHI database. An engineer working on a problem can look in the database of technical memos and see if that problem has been dealt with before. The goal of this database is to prevent duplication of work with all the attendent savings of time and money.

Manufacturing semiconductors, a core Intel business, involves many hazardous materials. Knowing the potential hazards of a chemical is critical for safety. DELPHI has on-line Material Safety Data Sheets (MSDS). These MSDS have data on the properties of materials used in chip fabrication.

DELPHI serves other information needs. Because many Intel employees are also Intel stockholders, DELPHI began offering stock price information. Opening and closing quotes and hourly updates between the two are available through DELPHI. DELPHI also provides files of general interest, such as standards documents and user guides. For such requirements as marketing and competitive intelligence, DELPHI provides subject specific news. Subjects such as Japanese semiconductor business and Intel related news are collected from news services, evaluated, and made available through DELPHI.

III. Connecting Services to Environments

Once DELPHI began offering services, the implementors of DELPHI needed to find a way to connect those services to the different environments throughout the company. Using an interface

accessible from virtual terminals and through electronic mail proved to be ideal for providing connectivity. This section describes how DELPHI can be accessed from the different network environments at Intel.

First , DELPHI applications had to be loaded on a computer. A spare VAX 8350 was located and christened DELPHI, after the ancient Greek oracle at Delphi. DELPHI runs the VMS operating system and can communicate using DECnet protocols. One of DELPHI's first applications, the bulletin board for Internet forums, was implemented using a public domain bulletin board application from MIT. BASIS from Information Dimensions, Inc., was chosen as a database management system for handling book, periodical, and technical memo information. Other information services, such as Intel stock price information, were developed locally and installed on DELPHI. DELPHI runs 24 hours a day, seven days a week in order to provide service to Intel employees in many different time zones.

Although information was placed on DELPHI, the problem of how to make that information accessible remained. One option is to develop client programs for each application on each computing platform in each environment. While this is technically feasible, the owners of DELPHI, the Library Systems Group (LSG), did not have enough staff to create, distribute, and maintain this software. A better solution was needed.

The most common network applications in Intel environments are virtual terminal/remote login, electronic mail, and file transfer/sharing. Since all of Intel's four environments implement some kind of virtual terminal protocol, a "login" interface to DELPHI's products was created. The login interface allows a user to log onto DELPHI using a special login ID that does not need a password. Once logged on with that ID, the user sees a menu of applications from which to choose. Figure 1 shows the first menu. The menu is made of simple text characters, with no elaborate graphics. This was done to make the menus usable from many machines as possible, because graphics control sequences often are vendor and protocol-specific.

Creating a menu interface for users is easy; the hard part is finding ways for users to reach that menu interface. Since DELPHI was a VAX running VMS, it could easily interface with the manufacturing environment using DEC's CTERM protocol. DEC's LAT protocol provides access from terminal servers at three different Intel sites. As a result, manufacturing had first access to DELPHI's information resources.

```
    ----------------
    DELPHI MAIN MENU
    ----------------

HELP  Help

B     Bulletin Board
D     Databases
M     MSDS Online
Q     Stock Quotes
S     Suggestion Box

LO    Logout

Selection  --->
```

Figure 1: DELPHI main menu

The other environments proved more difficult. LSG needed a simple way to connect DELPHI to the three remaining areas. TCP/IP based protocols are the de facto method of connecting computers from different vendors. TCP/IP protocols are not proprietary, and there are implementations on many different platforms. The DELPHI implementors decided to push TCP/IP as the connectivity solution.

The first step was to install TCP/IP on DELPHI. This was not difficult, as a number of TCP/IP implementations are available for VMS. This step provided immediate access to DELPHI for the engineering environment because that environment already used TCP/IP and telnet. Fortunately for the DELPHI implementors, BANYAN VINES has an option to have TCP/IP bundled with it. Parts of the administration world thus gained access at the same time.

Providing access to the IBM mainframe world was the last task. Although a DECnet/SNA gateway was available, it was rejected because special software would have to be run on DELPHI to make it work. The DELPHI system administrators did not want another piece of software to maintain. A TCP/IP implementation on MVS (an IBM operating system) would be a much better solution since no additional software would be needed. To achieve that end, DELPHI's implementors helped push for and realize TCP/IP in the business world, working with Intel's MIS organization. In a diverse network, interorganizational cooperation is critical toward connecting disparate environments.

While this work went on, the only network access for some users was and remains electronic mail. Virtual terminal capability, while widespread, is not universal within Intel. DELPHI thus had to communicate to several different dialects of electronic mail. It already uses MAIL11, the DECnet mail protocol, to connect to the manufacturing world. Additional software was purchased to implement SMTP, the Simple Mail Transport Protocol used in the engineering world. An SMTP/cc:Mail gateway was available to link to the administrative environment. DELPHI still needed links to the business mail system on mainframes and Banyan mail. To help connect these remaining areas, DELPHI's implementors participate in a corporate project to link Intel's mail systems. Working again with Intel's MIS group enabled DELPHI's implementors to leverage corporate resources in order to gain more connectivity.

Electronic mail delivers services in a variety of ways. Subject specific news is distributed through electronic mail. A mail server was installed to automate much of the process. Functioning like a BITNET LIST Server, DELPHI's mail server takes certain commands in a mail message and processes them. Users can subscribe and unsubscribe to subject specific news on their own. Internet forums are redistributed this way. In addition, files and documents of general interest can be retrieved through the mail server.

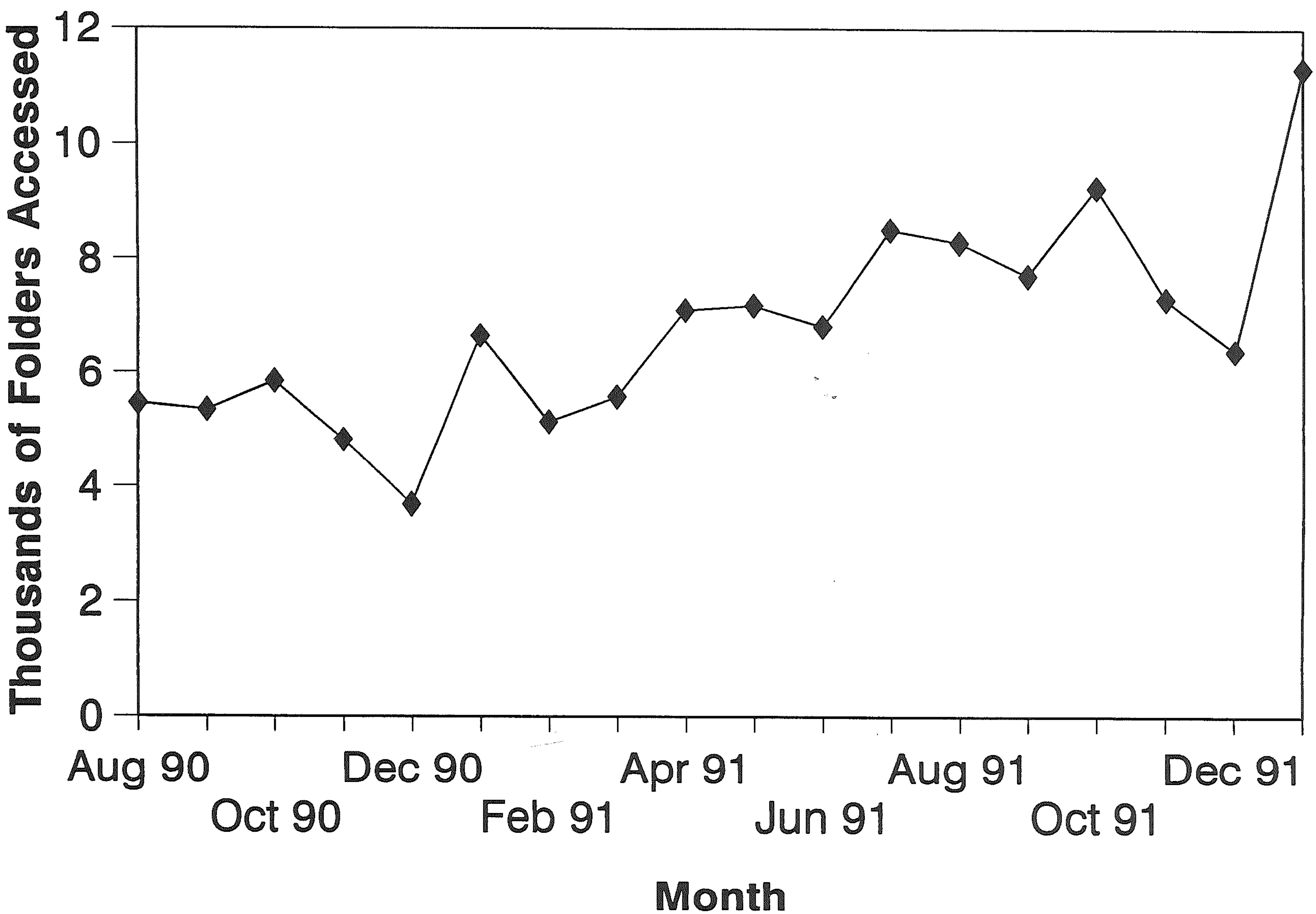
DELPHI's handling of stock information is a good example of how DELPHI provides specific information to different network environments. Many Intel employees are also Intel shareholders, and they have a strong interest in Intel's stock price. Since these shareholders reside in different environments, DELPHI provides different ways to get the stock price information. First, the DELPHI login menu includes a choice for obtaining Intel's stock price. Second, the stock price is also available in DELPHI's bulletin board. Third, if users wish to avoid logging onto DELPHI (or cannot), they can use DELPHI's mailserver to get the stock price. They also can use a special telnet server implemented on DELPHI or DECnet's copy facility to get price information.

IV. Experiences with DELPHI

Implementing DELPHI provided a mix of experiences. While DELPHI became popular and heavily used, success brought its own set of problems. Other problems were caused by the design of DELPHI applications. This chapter describes our experiences with running DELPHI.

DELPHI services proved to be popular -- too popular. As a result, the computer becomes bogged down and slow during working hours. All available login ports are used at times. An example of the use that DELPHI experiences is shown in figure 2. DELPHI's bulletin board use has climbed steadily since usage has been tracked. Other services, such as the databases, show similar growth. LSG's experience is that any CPU cycles saved by fine tuning DELPHI are consumed by users. Anecdotal evidence suggests that DELPHI would be used even more if it had more capacity.

Figure 2: Bulletin Usage



The availability of Intel stock information on DELPHI produced interesting usage patterns. On days when Intel stock was very active, DELPHI would be particularly slow as Intel employees check the changes to their net worth. While this indicates that DELPHI's information resources are being heavily utilized (a positive point), it also points out flaws in the DELPHI design.

The first mistake was putting all of DELPHI's applications on one machine. If one application slows down DELPHI, all of DELPHI's applications will be affected. Centralizing DELPHI, while making management easier, creates a single point of failure. Once DELPHI is down, all of its information resources are inaccessible. Part of this problem stemmed from the ad hoc way that DELPHI's applications evolved. New applications were developed, put on DELPHI, and added to its menu of applications without much planning and thought to long term strategy.

Another mistake was implementing DELPHI in a poor location within Intel's network. The system is not located toward the center of Intel's network. Instead, it is located toward the edge of the network. On average, network traffic must travel farther, making interactive network response poorer and making DELPHI's applications seem slower. Overseas sites suffer the most because their network connections are slower than US sites because of costs.

Finally, DELPHI lacks sufficient performance and network tools to find problems. Users might complain of "slowness," but without good network and system management tools, it is difficult to isolate the cause.

Despite these problems, DELPHI must be considered a success. Its services are reachable from all of Intel's environments. It is used heavily and often. User profiles reveal that it is used throughout the 24 hours a day that it is available. While DELPHI definitely has problems, lack of use and interest certainly is not one of them.

V. Future Plans for DELPHI

Plans for DELPHI involve building upon its successes and correcting its deficiencies. First, DELPHI will move to a bigger machine. Its CPU has become the limiting factor in its performance. The new DELPHI will be much more powerful, and it unlike the current model, will have an upgrade path. Speeding up applications and adding capacity for more users will generate even more usage.

Additional information and network resources are being considered for DELPHI. Important corporate databases, such as the company phone book and the Intel electronic mail directory could be placed on DELPHI. Access to the increasing number of information resources available on the Internet will be looked at.

Networking is another area that will be improved. Intel's network is being rearranged to move DELPHI much closer to the center of the network. This will reduce interactive network delays, and users should see a gain in responsiveness. As new network environments like Novell are introduced, ways must be found to provide connectivity. Much of DELPHI connectivity depends on the TCP/IP protocol suite, In the long term, this connectivity will migrate to the OSI protocol suite as standards evolve and are implemented.

Linking DELPHI's bulletin board system to the corporate Usenet News network will be examined. Doing this would make DELPHI's information resources available directly to many Intel employees without having them log into DELPHI. DELPHI would be less loaded, and fewer users would spend time navigating DELPHI's menus.

Finally, the long term strategy for DELPHI is being formed. The owners of DELPHI, together with other groups in Intel, are creating a strategy for an Intel information architecture. This strategic roadmap will where DELPHI fits into that architecture and how it will evolve. This should lead to planning and avoid some of the problems of ad hoc implementation.

Conclusion

DELPHI successfully provides an information resource in a multivendor, multiprotocol network environment. Some key techniques toward achieving this goal are encouraging the use of "open" and

widely implemented protocols, working with other corporate organizations to build connectivity, and creating application interfaces to nearly universal services like virtual terminal/remote login and electronic messaging. Most of the DELPHI's shortcomings are the result of its very success. With an increase in its capacity, network improvements, and the creation of a strategic roadmap, DELPHI promises to be a success for a long time to come.

REFERENCES

Comer, Douglas. Internetworking with TCP/IP: principles, protocols, and architecture. Englewood Cliffs: Prentice-Hall, 1988.

Digital Equipment Corporation Guide to DECnet-VAX Networking. Maynard: Digital Equipment Corporation, 1988.

Digital Equipment Corporation VMS Networking Manual. Maynard: Digital Equipment Corporation, 1988.

Quarterman, John S. The Matrix: Computer Networks and Conferencing Systems Worldwide. Bedford: Digital Press, 1990.

Rose, Marshall T. The Open Book: A Practical Perspective on OSI. Englewood Cliffs: Prentice-Hall, 1990.

Weber, Robert. Intel Library Technical Architecture and Strategic Assessment. Consultation Report for Intel, 1991.

The Campus Marathon: The Transplant of the New Connecated Heart

Anastassia Khouri St.Pierre
Systems Librarian
McGill University

In the past two decades library automation has ceased to be an option and has become a necessity. The implementation of integrated systems moved slowly but surely. Some institutions moved to their second and a few to third generation systems with a diversification of access to collections, information, and services. With the continuing evolution and sophistication of communication technologies, academic libraraies are moving slowly again from library networks to campus, national, and international networks. This shift is offering the academic library the opportunity to redefine its technical options and services. This paper will look at the impact of the new networking role of the academic library both on library services and on its mandate as an information provider to support teaching and research. Secondly, I will discuss the relations between the academic library and the campus computing centre. I will argue that an innovative cooperative partnership is required. The academic library has to plan its services and initiatives according to the technical capabilities of the campus computing centre. The mosaic of services offered by the academic library will be "dictated" by the technical sophistication of the computing centre and its capability to support adequately a complex communication infrastructure. Expansion of the role of the library may even require that its partnership with the computing centre extend to joint requests for additional funds. The mosaic of library services will be unique to each campus. For the next decade, I foresee that we will use a completely new set of criteria to evaluate academic libraries. In that respect, what will be the role of the computing centre? The libraries were always cited as the heart of the University. Are we going to transplant the old heart for a new connected one?

THE ROLE OF PUBLIC LIBRARIES IN THE USE OF INTERNET/NREN INFORMATION SERVICES: PRELIMINARY FINDINGS

By

Charles R. McClure (cmcclure@suvm.acs.syr.edu)
Joe Ryan (joryan@suvm.acs.syr.edu)
Diana Lauterbach (dmlauter@suvm.acs.syr.edu)

School of Information Studies
Syracuse University
Syracuse, NY 13244-4100

ABSTRACT

This paper provides preliminary results from a study funded by OCLC to investigate the role of public libraries in developing and exploiting the next generation of national networks as presently embodied in the Internet/NREN. The results from the study are intended to identify a range of roles, services, and responsibilities for the public library community as it becomes an "electronic intermediary." Moreover, the paper identifies key policy issues and offer some preliminary recommendations to enable public libraries to better transition to, and operate in, the future national networking environment.

THE ROLE OF PUBLIC LIBRARIES IN THE USE OF INTERNET/NREN INFORMATION SERVICES: PRELIMINARY FINDINGS

At the recent White House Conference on Library and Information Services (WHCLIS), futurist Clement Bezold offered possible scenarios for the future of the public library: libraries fade away, libraries in cyberspace, and post industrial libraries in the search for a more just society (Bezold, 1991). While clearly there are other possible scenarios, the question of how public libraries will evolve in the electronic networked environment remains a largely unaddressed and unanswered question. In the age of communications, will the public library survive? Or will it be killed by technology? With fiber optic networks that can deliver library materials directly to the user from computerized data banks, is there any need for the library function (Wicklein, 1983, p. 2)?

These questions, have been considered for some time, but have gained in importance as Wicklein's predicted future becomes reality for today's public library. How will the opportunities and challenges posed by newly emerging networked, information resources and services be integrated into the traditional areas of public library activity? How should public libraries use the developing electronic networks to assist public libraries in this new environment?

Simply stated, the problem is that public libraries are likely to be the most neglected by national electronic network planners. Yet public libraries have the potential to generate some of the most innovative educational uses of the network for the widest range of individuals. But public libraries may have the greatest difficulty adapting to the new electronic networks. Early advanced planning and needs assessment can increase the integration of networked resources and services into public library practice.

This paper provides preliminary findings from a study sponsored by OCLC to explore possible roles for the public library in the evolving networked environment. Additional findings and issues based on yet to be completed data gathering and analysis will be provided on-site at the ASIS midyear conference. But clearly:

- There is much work to be done in increasing the awareness of the public library community about electronic networking.
- Network planners, policy makers, and public libraries have yet to fully understand a range of issues affecting the public library's involvement in the networked environment; and,
- Specific roles, services, and activities for the public library in the networked environment have yet to be identified.

How will public libraries evolve, survive, thrive?

BACKGROUND

On December 9, 1991, President Bush signed into law the High Performance Computing Act of 1991. In addition to mandating research and development related to high performance computing, the Act authorized the establishment of the National Research and Education Network (NREN) and became Public Law 102-194. The process by which the bill was introduced, debated, revised, re-introduced, was the subject of hearings, and lobbied during the past three years was tortuous. But the bill did become law (McClure, Bishop, Doty, and Rosenbaum, 1991).

The act will dramatically upgrade and expand the existing information resources and services available on the existing Internet network. Lynch and Preston (1990, pp. 280-281) describe the Internet as follows:

> In effect, then, the Internet includes hundreds of institutional or corporate "local-area" networks some of which contain thousands of computers), a series of NSF [National Science Foundation] regional networks, the NSF backbone (which is the primary transcontinental traffic path), MILNET [military], and a range of other agency-specific or experimental networks. The Internet provides connectivity among perhaps half a million computers and over a million people, most of them within the research and higher education community. The system is also linked internationally to networks in Europe, Japan, and Australia. Electronic mail can flow between the Internet, BITNET [a popular cooperative research and education network], and commercial services such as Compuserve and MCIMAIL, further increasing the scope of communications available to the Internet user community.

The Internet, in turn, can be viewed as a prototype for the U.S. federally funded, National Research and Education Network (NREN). According to Bishop (1990) the legislation will:

- Establish a Federal High Performance Computing Program in which science agencies and national libraries will fund and conduct research, and develop technologies and resources, appropriate for the NREN.
- Mandate the creation of the NREN -- to link over 1,000 Federal and industrial laboratories, educational institutions, libraries, and other facilities -- over the next five years.
- Promote the development of a number of electronic information resources and services on the NREN, such as directories of users and databases, electronic journals and books, access to computerized research facilities, tools and databases, access to commercial information resources and services, and user support and training.

Senator Gore has described the NREN as a "information superhighway." Senator Hollings, a key supporter of the legislation suggested that the NREN "could become the most powerful teaching tool ever built" (Hollings, 1990, p. S18114). $638 million

has been budgeted for this initiative by the federal administration for the present fiscal year (Office of Science and Technology Policy, 1991, p. 2).

A new generation of electronic networking is poised to begin. The possible network applications range from electronic mail, list-serves, file transfers, remote access to computing, and electronic reference services, and uses just beginning to be contemplated by library community. As we move into the next generation of the Internet/NREN, other networks uses become not only possible but a competitive necessity.

Public libraries, despite extremely limited funding, have been key innovative players in the development and use of the educational components of electronic networks. Examples include:

- The use of telefacsimile for document delivery and communication (Jensen, 1988); videotext and teletext (Appleman, 1984; Pollard, 1983) including OCLC's Project 2000; cable television services (Chepesiuk, 1985); community satellite dishes (Amdursky, 1985); distance learning (Burge, et al,1989); rural library - college links (Vasey, 1989); and improved service to the physically handicapped (Jahoda & Needham, 1980);
- Community databases (Ahtola, 1989) including emergency services (Magrath & Dowlin, Spring 1987), events calendar, government agency directories and access, career services and travel information (Malyshev, 1988; Dowlin, 1984); electronic bulletin boards (Dewey, 1984; Dewey, et al., 1985; LaRue, 1986); and, electronic mail (Kemper, 1988).

The Internet/NREN offers a context for the development of library services and the provision of resources which have yet to be investigated.

Will the mission of public library service remain the same? That "library resources be equally available to all citizens of the community, and that the collections attempt to represent the widest possible number of viewpoints" (Dowlin, 1984, p. 24)? Will the library function remain the same: "an institution guided by trained intelligence that serves as an editor and consultant...for the public concerning the information it needs (Wicklein, 1983, p. 7)? Will this function be one "...that we in the general public must for the most part delegate, if we are to make sense out of the vast amount of material available to us" (Wicklein, 1983, p. 8)?

Public libraries may find that their role in the community may change significantly as a result of access to the NREN. In previous work on planning, McClure et. al., (1987) developed eight service roles from which public libraries may choose to meet community needs: community activities center, community information center, formal education support center, independent learning center, popular materials library, preschooler's door-to-learning, reference library and research center. New visions and service roles will need to be developed as a result of Internet/NREN use.

Another concern is the finding from a recent study that "key players in the Federal government have given little attention to how the library community could be involved [in the NREN] (McClure, et. al., 1990a, p. 30). Findings from that study suggest that the library community, in general, and public libraries more particularly, have no clear sense of their role in the Internet/NREN environment. The proposed Internet/NREN will present great challenges and opportunities for libraries. But how, exactly, the library community in general, and public libraries in particular will make use of the Internet/NREN is unclear.

STUDY METHODOLOGY

This study seeks to provide a description and assessment of key issues affecting public library roles in the use of non-bibliographic, Internet/NREN information services. The study addresses, in an exploratory fashion, topics such as:

- How knowledgeable is public library leadership about present developments in the national electronic networks?
- What are the innovative ways that public libraries are presently using electronic networks (excluding bibliographic retrieval and location)?
- How are public libraries integrating networked information resources into their organization's service delivery?
- What new techniques are public libraries employing to improve organizational productivity and effectiveness using networks.
- What service roles might be developed for public libraries with the advent of the next generation of electronic networks?
- What Federal government information sources and services would public libraries like to access on the Internet/NREN?
- What future network services do public library leaders wish to see? Which should be adopted first?
- What barriers are public libraries likely to face when adopting the next generation of network technology?
- What specific steps are presently being taken to position public libraries to take advantage of the networked environment?
- What are the implications of public libraries' use of the Internet/NREN for OCLC?

Addressing such questions will greatly assist public librarians in using, managing, and adopting new roles as they move into the next generation of electronic networks.

The study focuses on public libraries because they have been the most neglected by national electronic network planners. Yet public libraries could generate important and innovative educational uses of the Internet/NREN. Moreover public libraries may have the greatest difficulty exploiting the new networked environment due to a number of organizational and managerial constraints. Thus,

two key groups of participants comprise the study population:

- Public library leaders -- targeted because this group is likely to be most aware and most in need of information about the Internet/NREN environment.
- Practicing public librarian middle managers with network familiarity -- chosen because their knowledge base determines what is practical to accomplish today and tomorrow.

As the study progresses, additional participation from other stakeholders will be obtained.

This exploratory investigation is based on a two-phased approach: (1) obtaining descriptive information regarding public library Internet/NREN uses, futures and potential impacts, and, (2) analyzing that descriptive information in light of various policy issues. The study relies on quantitative and qualitative methods as well as a range of data collection strategies. Findings reported in this paper are based on literature analysis, focus group sessions, and individual interviews with public library leaders and managers.

Much of the data collection relies on focus groups which are particularly useful in social science research that is exploratory and aimed at the generation of hypotheses and research questions (Krueger, 1988). This technique has been previously used, most successfully, by the researchers studying scientific communication and electronic networking.

Additional data collection methods are nearing completion at the present time. These include:

- Computer assisted content analysis of the results of the focus groups and interviews conducted to date.
- Analysis of focus group participant profile data.
- Analysis of a survey questionnaire administered to targeted public librarian samples of convenience attending electronic network sessions at national conferences.
- A case study involving public library participation in electronic networks.

These multiple data collection techniques will assist the researchers in examining the topic from a range of perspectives and increase the likelihood of collecting valid and reliable data.

KEY ISSUES AND FINDINGS

The following key issues and findings that affect the development of public libraries in the networked environment can be reported based on the literature review, interviews, and focus group sessions conducted to date. We anticipate that

some modification of these findings will occur and additional issues will be identified as the additional data collection efforts described above are competed.

General Enthusiasm for National Networking

Public librarians are enthusiastic about national networking as represented in the literature or in speeches they have heard. But while enthusiastic, they raise concerns about what, specifically, national networking has to offer the public library setting. For example, one person commented that she never has time to just sit in her office and use any system for any period of time without interruptions--how would she have the time just to do networking on top of all her other job responsibilities? While the concept of the NREN and remote access to information looks intriguing and may have the potential to significantly change public librarianship, what "national electronic networking" actually is remains pretty vague to most public librarians.

Awareness

Many of the participants commented on the limited awareness of networking issues that public librarians typically had. They doubted if the vast majority of public librarians knew about the NREN, what it was, how it worked, and the information resources/services that it carried. They thought that inadequate attention had been given to NREN issues in public library literature. One person commented that she had thought it had only to do with research and academics and did not realize that other applications might be useful for public libraries.

Librarians noted that little discussion of the NREN or national networking topics and issues occurred in their local libraries, or library association meetings. They rarely discussed such issues among themselves (although one said that they certainly would be now after having participating in a focus group session). Librarians felt that the profession as a whole had little awareness of the key issues or topics related to the NREN and national networking.

Risks Associated with NREN Involvement

Some participants mentioned the risk-taking aspect of utilizing "unproven" new technologies and wondered if "the train had left the station" or "had it not yet arrived?" As a director described the situation, she wanted to be "out front in the use of new technologies, but safe enough that they would not change out from under her." There was general agreement that separately, public libraries did not have the resources to take on such risks associated with developing the uses and applications of networking. They needed someone else to develop, implement, and test applications FIRST. There is no slack in current public library budgets to try something just because it may be a good idea.

An interesting aspect of this issue was the consensus on the need for public sector entrepreneurial perspectives in the public library. When asked who, exactly, should be taking these risks they felt that someone in public librarianship should, and probably someone in the public sector because it was unlikely that others in the private or Federal government sectors would take on such a responsibility. There was general agreement that it was a very difficult time to be taking "technology risks" given the existing economic climate for many public libraries.

Barriers to Network Use

The group of traditional barriers mitigating against the development of networking in public libraries includes: limited knowledge about the Internet, inadequate equipment, and limited staff knowledge in the use of computers and telecommunications; confusing and contradictory information about how to connect to the network; no "systems" people to implement the network in their libraries; and no time to commit to such activities. In addition, some public librarians are unconvinced that there was public library "stuff" useful to them via the Internet. They recognized that the network user should have a range of skills and knowledge -- especially in commands and systems protocols -- which they did not have and were unlikely to get in the near future.

There also was the perception that the organization of information and resources on the Network was a "mess" and they saw that as a serious barrier in their effective use of the Internet: "How can I use the information if I don't know what's out there or can't locate it?"

A number of responses showed special concern about how public librarians would be re-educated to meet the challenges of operating in the networked environment. All agreed that the host libraries had to do a better job of developing continuing education programs, that professional associations had to support such efforts (perhaps with post-MLS certification requirements), and that sabbaticals or support for public librarians to leave the job situation to be re-educated were needed.

Connecting to the Internet

Although it was mentioned in the context of a barrier, the issues of how exactly one gets connected to the Internet, how that connection is made available throughout the library system, and the costs associated with this connection process were raised repeatedly. Public librarians want a step-by-step listing of what exactly they had to do in order to get connected and use the Internet. They wanted to know what the connection costs were, they wanted to know who best to contact to get the connection, and they wanted to know NOW. Such information is not available to most public librarians.

In a number of different conversations with different librarians in different parts of the country, the theme of poor technical information and instructions for connecting with the Internet/NREN was consistent. One respondent commented that she had talked to her local bibliographic network, a regional network, a private vendor, and OCLC about how to "get connected." In each instance she received conflicting information and wide ranging estimates of the time, expenses, and level of effort that would be needed to connect to the national network.

Access to Networked Information

Some respondents thought that having public access terminals to the Network in the public library could be a good idea. This would support the role of the library protecting those with less resources and computer literacy to still have a "safety net" where they could get on the Internet. There was general consensus that the public had a "right" to the Network and it would be good for the public library to be an intermediary to provide this access. There was some split opinions about the increasing use of home modems to access either the library or the Internet directly.

One participant, however, immediately recognized direct access to the Network without going through the library as a significant threat to the public library: "if all this information is available directly to patrons and they do not have to come to the library to get it, why will they support the public library?" Additional discussion took place on this topic, but it was clear that a number of librarians began, for the first time, to consider the Internet/NREN as a threat rather than an opportunity for public libraries.

Public Library Information and Services on the Internet

A common question raised by the public librarians is that they wanted to know what exactly there was on the Internet that might be useful for them NOW. When the investigators listed a number of "typical" information services and resources currently available, they clearly were unimpressed. The sense was that Internet services and resources needed to be developed and designed specifically for the public library community. They suggested that real "down-to-earth" information services and products would be necessary if John Q. Public was to use public library to access the Internet. The kind of services they suggested were:

- Full text, color children's books on the network
- Practical listservs such as recipes-l; autorepair-l; homework tips-l; or crafts-l
- Community-based information services in health care, community activities, and unique local resources
- "Job-net"
- Dissemination and access services linked directly to the, responsibilities of local governmental units in the city or county

- Remote access to library reference and referral services
- Support for local schools and specific instructional and curricular activities
- Making government databases accessible to the public via the Internet rather than having to go through existing vendors.

But overall, it was difficult for the librarians to describe specific types of public library services that could be offered using the Network. As one person commented: "we are real concrete people, what exactly does this network look like and how can I use it? Until I figure out how I can use it I can't visualize it."

Role of Professional Associations

One person commented that the Public Library Association (PLA) board had recently discussed the role of the public library in the Internet/NREN (Summer, 1991) but not much had come from it. She attributed this to the fact that the Network was too vague to understand at this point: "Frankly, the board can't figure out what to do with this issue." A public library branch manager pointed out that the people who knew most about the Network were likely to be junior staff and not the library directors or members of the professional association boards. Thus, she was concerned that change would occur very slowly since the people with the most power know the least about what needed to be done to exploit the Network.

There was wide agreement that if ever there was a time for state libraries to take a leadership stance in the use of the Internet for public libraries, it was now. A majority believed that the locus for coordination of statewide development public library development of the NREN should be the state library and that they needed to coordinate that effort with the State Education Department and local governmental units. There was also agreement that it was unlikely that the state libraries were up to the challenge given the financial difficulties many states are experiencing.

Committing Resources for Network Access/Use

Participants made it clear that they all had tight budgets and now was a very difficult time to come up with resources to support a new initiative such as access to the Network. This was all the more problematical since nobody really knew how much it would cost and what exactly the benefits might be for the library and the community. In fact, the sense was that UNTIL a better understanding of what the costs were and what benefits would be obtained (for both the library and its patrons), resources would NOT be committed to this initiative. Some of the librarians suggested that it would be extremely useful to develop models, or typologies of possible costs for the public library to get involved in the NREN at a range of levels of efforts and services provision.

Getting Involved

The librarians offered a number of specific recommendations for how the public library community could become a player in access to and provision of networked information services:

- Develop a model Internet-connected public library and show others what CAN be done and what the library can do with Internet based information services
- Develop arrangements where public libraries with unique resources in one location make those resources available to other libraries via the Internet
- Educate state library and association leaders as to the key issues regarding public library use of the Network
- Initiate a massive program to increase the awareness of public librarians regarding this issue and then start a re-education program nation-wide
- Demonstrate to local governing bodies what access to the Internet might do for them, locally.

Overall, it seems that public librarians were very interested in becoming "networked" and that they wanted to be part of the NREN. Moreover, they saw a potential to provide information services to target groups that might not otherwise have access to electronic information. But they did not know what to do to get started, how to start-up the connections, what to do once they got the connections, and how to convince their funding bodies that re-allocation of resource to Networked activities was "worth it."

RECOMMENDATIONS

While data collection and analysis is still in process, it is apparent that a number of preliminary recommendations can be offered to assist the public library move into the evolving networked environment.

Need for Good Examples

An ongoing theme in the discussions was the need for some good examples of good examples of what to use the networked environment for in a public library. More than once people asked why didn't we have a video tape of using the Network in a public library context rather than an academic library context--as had been done with the "Beyond the Walls" video tape (NYSERNet, 1991). The unsaid implication was "there probably really isn't much you can do with the network in a public setting, is there?"

In short, we need to develop a concrete set of examples of what to do with the Network in a public library context. This set of examples might come from producing a video tape or it might be in developing a "showcase" public library in its use of networking that others could see in a "hands-on" context. For many

public librarians, something concrete and real was needed for them to appreciate the use and applications of the NREN.

Need for Education

Assuming that we can resolve the awareness problem and increase the public library's knowledge about the importance of networking issues, there are still massive re-education problems to be addressed. Public librarians recognized the need for them, personally, to be re-educated but they had a range of problems and fears regarding the process. A program of educational opportunities related to the Internet/NREN need to be developed with cooperation among the libraries, the professional associations, the library schools, network providers, and federal and state governments. Additionally, mechanisms for providing incentives and rewards for librarians to participate in such programs are essential.

Leadership

Currently, there is a leadership void in addressing the role of the public library in a nationally networked environment. There must be leadership in the profession to confront applications and uses of the Internet for public libraries. This is not seen, currently, as a key issue at most public libraries, PLA, or at the state libraries. Individual library directors may recognize its importance but do not know what to do about Interneting in THEIR library. Who will come forward to provide the leadership necessary to connect public libraries to the Internet and show them how to use it to meet community information resources? Some participants worried "are we up to this challenge? I haven't recovered from the preceding challenges I have had to deal with on this job!" The leadership issue and mounting support within the public library community to deal with national networking is critical.

Clarify Connection Confusion

There currently is great confusion about how, exactly, a public library can get connected to the Internet. Apparently there are very few vendors that are concentrating on the public library market for connection to the Internet. Public libraries do not know who to go to for information regarding connection, costs, applications, and training.

Part of the confusion stems from connecting to the Internet/NREN being primarily a local issue. The way in which a library in Georgia might get connected could vary considerably from how a library in California might get connected. This variance stems from the ease of access to regional or mid-level networks, the type of cables available to the library for connection, and the support that the local regional (or other provider) might be able to offer. This confusion adds to the mysterious nature of how the public library might connect to and use the Internet.

Models for Network Involvement

Public libraries need a set of possible models for how they might get involved in the Internet, what costs might be associated with what models, and what types of services and benefits might be realized from a particular model. Factors to consider in the development of such models include:

- Size of the library
- Organizational structure of the library and how it reports to its governing body
- Nature of the library clientele and the range of services to be provided
- Level of effort that can be committed by the library to networking
- Staff knowledge and interest in Internet/NREN services/involvement
- Existing technology infrastructure.

In fact, it might be that there are different levels of effort associated with the various models. This would allow the public library some flexibility in how it might develop its network participation.

Impact of Planning and Roles Setting Manual

It appears that <u>Planning and Role Setting for Public Libraries</u> (McClure, et. al, 1987) could be a serious inhibitor to developing the networked public library. The roles in the manual are very traditional and do not address activities associated with electronic provision of information resources. Moreover, many public libraries (including those who participated in this focus group) have used the Planning Manual and expect that a service role (e.g., the Networked Public Library) should be developed and added to the "acceptable list" of public library roles prior to moving in this area. A range of new service roles and vision statements for public libraries in the networked information age are needed.

THE ROLE OF PUBLIC LIBRARIES IN THE NATIONAL NETWORK

Perhaps the single most important factor that is needed for greater public library involvement in the Internet/NREN is vision. A vision statement is a description of a possible future state or set of functions for the library. Vision statement development requires librarians to make explicit their assumptions about the future and to envision a future state of the organization in light of these assumptions and in light of organizational goals and resources.

A primary purpose of vision statement development is to define and describe visions of what the library might be in the future networked environment. In terms of strategic planning, the library can develop a range of possible visions, identify those that are most important and which would benefit the library and its clientele the most, and then take appropriate steps to insure that the vision evolves as

defined. As such a vision statement provides a target at which the library can shoot, a vision of what it would like to become, and suggestions for the resources that will be needed. Currently, there is little vision of what the public library might be in the nationally networked environment.

A key notion of vision statements is the idea of taking responsibility for the development of a library's future and not letting that future occur by happenstance. For the public library community to take charge of its future, attention must be given to:

- Developing national spokespersons for articulating the role and responsibilities of public libraries in the Internet/NREN environment
- Affecting national policies on how the Internet/NREN will be funded, used, and integrated into the public library community
- Increasing, exponentially, the awareness of and knowledge about the Internet/NREN in the public library community.

As this research project continues, suggestions and strategies will be offered to address these, and related issues.

The fabric of our society continues to change as a result of the evolution of the national network. The library community, in general, and the public library community more specifically, must change as well. The evolving role for the public library in the networked environment can be the traditional safety net role that insures access to the network by all citizens. But its role can also be "electronic navigator and intermediary," it can be "provider of electronic information to remote users," and "switching station" among the possible electronic information resources and services. But these roles must be created; visions for these roles are needed now; and immediate involvement in the design and structure of the Internet/NREN are needed to insure that the public library is a key player and stakeholder in the evolving national networked information society.

REFERENCES

Ahtola, A. Anneli. (Fall 1989). In-house databases: An opportunity for progressive libraries. RQ, 29 (1), pp. 36-47.

Amdursky, Saul J. (1985, November 15). Dishing it out: Satellite services in public libraries. Library journal, 110 (19), pp. 49-51.

Appleman, Merrie. (1984). Videotex: Options for libraries. Report Number Five of the Electronic Text Report Series. San Diego, CA: San Diego State Univ., Center for Communications 1984. Available from ERIC, ED 259722.

Bezold, Clement. (1991). Libraries in the 21st century: Alternative futures for the great debate. Paper presented at the White House Conference on Libraries and Information Services, July 12, 1991, Washington, DC.

Bishop, Ann. (1990, November). The National Research and Education Network (NREN): Promise of a new information environment. Syracuse, NY: ERIC Clearinghouse. ERIC Digest. EDO-IR-90-4.

Burge, Elizabeth J. et al. (1989). Interactive libraries; Dimensions of interdependence. Paper presented at the International Council for Distance Education Conference (Cambridge, England, September 19-22, 1989). Available from ERIC, ED311907 (see also ED311902).

Chepesiuk, Ron. (1985, May). Information around the clock: Atlanta's channel 16. Wilson library bulletin. 59 (9), pp. 597-599.

Dewey, Patrick R. (1984, Spring). The Electronic bulletin board arrives at the public library: The North-Pulaski library prototype. Library hi tech, 1 (4), pp. 13-17.

Dewey, Patrick R., et al. (1985, Nov.-Dec.). Library use of electronic bulletin board systems. Library software review, 4 (6), pp. 351-361.

Dougherty, Richard M. (1991, February). An ideal win-win situation: The national electronic highway. American libraries, 22 (2), p. 182.

Dowlin, Kenneth E. (1984). Electronic library. NY: Neal-Schuman.
Harriman, Sigrid G. (Ed.). (1986). Toward a common vision in library networking. Proceedings of the Library of Congress Network Advisory Committee Meeting. Washington, D.C.: Library of Congress, December 9-11, 1985. Network Planning Paper No. 13. Available from ERIC, ED286542.

Hollings, Senator Ernest. (1990, October 24). Congressional debate on the National High Performance Computing bill. Congressional record. p. S18114.

Jahoda, Gerald, and Needham, William L. (1980). The current state of public library service to physically handicapped persons. Final Report. Tallahassee, FL: Florida State University, School of Library Science. Available from ERIC, ED313051.

Jensen, Jan et al. (1988). A.S.A.P. (As Soon As Possible): A multitype library fax consortium. Denver CO: Jefferson County School District R-1, 1988. Available from ERIC, ED 319389.

Kemper, Marlyn. (1988, November). Emerging technologies: A roadmap for librarians. School Library Journal, 35 (3), pp. 36-41.

Krueger, Richard A. (1988). Focus groups: A practical guide for applied research. Beverly Hills, CA: Sage Publications.

LaRue, James. (1986, June). Sending them a message: Electronic bulletin boards. Wilson library bulletin, 60 (10), pp. 30-33.

Lynch, Clifford A. & Preston, Celia M. (1990). "Internet access to information Resources." In: Williams, Martha E. editor. Annual review of information science and technology. New York: Elsevier, 25, pp. 263-312.

Magrath, Lynn L. & Dowlin, Kenneth E. (1987, Spring). "The potential for development of a clearinghouse for emergency information in the public library." Special libraries, 78 (2), pp. 131-135.

Malyshev, Nina Alexis. (1988). Concept and reality: Managing Pikes Peak library district's community resource and information system. Reference services review, 16 (4), pp. 7-12.

McClure, Charles R., Owens, Amy, Zweizig, Douglas L., Lynch, Mary Jo, and Van House, Nancy A. (1987). Planning and role setting of public libraries: A manual of operations and procedures. Chicago: American Library Association.

McClure, Charles R., Bishop, Ann P., Doty, Philllip, and Rosenbaum, Howard. (1991). The National Research and Education Network (NREN): Research and policy perspectives. Norwood, NJ: Ablex.

McClure, Charles R., Bishop, Ann, Doty, Philip, and Rosenbaum, Howard. (1990a). Realizing the promise of the NREN: Social and behavioral considerations. In Carol Parkhurst (Ed.), Library perspectives on the NREN. Chicago: American Library Association.

McClure, Charles R., Bishop, Ann, Doty, Philip. (1990b). Impact of electronic networks on scholarly communication and the research Ppocess: Final report. Washington DC: U.S. Congress, Office of Technology Assessment.

NYSERNet. (1991). Beyond the walls: The world of networkediInformation, an instructional workshop package. Syracuse, NY: NYSERNet, Inc. [111 College Place, Rm. 3-211, Syracuse, NY 13244].

Office of Science and Technology Policy. Federal Coordination Council for Science, Engineering, and Technology. Committee on Physical, Mathematical, and Engineering Sciences. (1991). Grand challenges: High performance computing and communications. Washington, D.C.: National Science Foundation.

Pollard, Richard. (1983). Videotex in libraries: An assessment of the British experience and directions for the future. Paper presented at the Mid-Year Meeting of the American Society of Information Science (Lexington, KY, May 22-25, 1983). Available from ERIC, ED237093.

Vasey, Bill. (1989). The intermountain community learning information services project: A participant's view. Rural libraries, 9 (2), pp. 105-107.

Wicklein, John. (1983). Will the new technologies kill the public library? Keynote speech at the LITA conference: "Information and Technology: At the Crossroads. September 20, 1983 in Baltimore, MD. Available from ERIC, ED237115.

CHARACTERIZATION OF INFORMATION RETRIEVAL SYSTEMS IN A NETWORK ENVIRONMENT[1]

Michael A. Shepherd and Carolyn Watters
Computing Science Division
Department of Mathematics, Statistics & Computing Science
Dalhousie University
Halifax, Nova Scotia, Canada B3H 3J5

ABSTRACT

A major challenge for information retrieval systems in a network environment is dealing with the lack of a global view of the multiple heterogeneous databases available for simultaneous access. In order to meet this challenge, it will be necessary to develop new methods of accessing and presenting information which are consistent with a user paradigm based on the needs of individual users. It would be helpful in developing such methods if there were a framework for viewing the state of the art, serving as a reference point for future and unforseen developments, and highlighting areas where creative thought and development are needed. Such a framework might be provided by the characterization, *from a user's perspective*, of information retrieval systems in a network environment along two axis; query and presentation. These axis reflect the ability of users to impose views within such information retrieval systems to best meet their individual requirements. The "query" axis reflects the ability of users to impose views at the time of query. The "presentation" axis reflects the ability of users to impose views on the retrieved information for presentation.

1. Introduction

In 1966, Flynn [1] categorized computer architectures along two axis, one axis indicating whether the architecture executed a single instruction at a time or multiple instructions in parallel and the second axis indicating whether the architecture processed a single data stream at a time or multiple data streams in parallel. This categorization provided a framework for the many initiatives in this area. We now find ourselves in a situation with respect to information retrieval in a network environment which is similar to the situation with architectures prior to Flynn's work, i.e., we have many new initiatives but no framework for these initiatives. Although taxonomies have been suggested for multi-DBMS and federated database systems [2], these tend to be based on database issues and are not entirely appropriate for the characterization of information retrieval (IR) systems *from a user's perspective.*

The network environment provides users with access to a wide spectrum of data and systems. The challenge for information retrieval systems in such an environment is in dealing with the lack of a global view of the multiple heterogeneous databases available for simultaneous access. In order to meet this challenge, it is necessary to develop new ways of accessing and presenting information within this environment consistent with a user paradigm based on the needs of individual users. A framework for these initiatives is important for viewing the state of the art, serving as a reference point for future and unforseen developments, and highlighting areas where creative thought and development are needed. It is useful to be able to qualify a system by its relationship to other existing or proposed systems by

[1]This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada, operating grant OGP0009163.

statements such as "X is an information retrieval system of type Y" with respect to one or more parameters of the given system. In the same way, one can reduce comparisons of "apples and oranges", in that comparisons can be restricted to systems of the same type. Finally, proposals for new systems can be made relative to the parameters of the taxonomy and thus enable more precise statements of differences from current systems of similar type.

This paper attempts to provide such a framework by characterizing information retrieval systems, from a user perspective, in a network environment along two axis; query and presentation. These axis reflect the ability of information systems to deal with the lack of a global view of the multiple heterogeneous databases available for simultaneous access. These axis reflect also the growing awareness of the need for "... individualization of information access." [3]. This means that in a network environment with shared access to multiple heterogeneous databases, users can impose their own views on the data in order to meet their individual requirements. Thus, in a network environment user view capabilities would include a single view of one or more data sets, multiple views of the same data set, and multiple views of different data sets. The ability to impose such individual user views in a network environment is reflected in the query and in the presentation capabilities of the systems.

The "query" axis reflects the ability of users to impose views at the time of query. This is really a function of how tightly the retrieval command language is coupled to the database structure. For example, the user may have access to a single set of commands designed to access a single database or multiple homogeneous databases. In this instance, the command set is tightly coupled to the database structure and the user should have the capability of imposing a view through appropriate structuring of the query. If the user has access to a single command set that is translated into multiple target command sets, then the initial command set is probably less tightly coupled to the database structures and the user will not have the same capability of imposing a view.

The "presentation" axis reflects the ability of users to have the retrieved information presented in such a way as to best meet their individual requirements. If the user has access to a single system, then a view can be imposed on the retrieved information for the purpose of presentation. If, however, the retrieved information is from multiple heterogeneous databases then it is much more difficult to impose a particular user view.

The remainder of this paper is organized as follows: Section 2 introduces information retrieval in a network environment, and Section 3 discusses each of the axis of characterization. In Section 4, various systems are characterized using these axis. Section 5, the conclusions, identifies areas where more or new research might be carried out based on this system of categorization.

2. Information Retrieval in a Network Environment

One can argue that information retrieval systems have been functioning within a network environment since the first time a user interacted with software on his local computer to access an online system on a remote computer. Since that time, we have seen a tremendous growth in the number of available databases, the distribution of those databases, and the functionality of network technology.

Such growth has lead, inevitably, to the development of standards for information retrieval in a network environment. Lynch notes [4] that information retrieval applications in a network environment fit the recognized client-server model, i.e., the user's machine is the client requesting services from a remote system. He then shows that there is a good match in functionality between general information retrieval applications and the ANSI/NISO Z39.50 [5] and the ISO Search and Retrieve [6] protocols for transmitting and managing queries and results. Although Lynch does not discuss simultaneous access to multiple servers, he does allow for a server to contain more than one

named information resource. In any event, an information retrieval network application should provide location transparency, i.e., all data should appear to the user to be at the local site, even though it may be at one or more remote sites. However, access to a wide range of servers by a single client, no matter if the servers are accessed simultaneously or one at a time, presents challenging user interface problems. These problems include view definition over various types of information resources, which is reflected in the query capabilites and in the presentation of information retrieved.

The direction of current work, discussed in Section 4.3, is to develop systems that make it appear that many different information retrieval systems are performing as a single virtual system and that all of their databases are a single virtual database. Such systems are called interoperable systems or multidatabase systems [7]. One must be careful, however, to distinguish between a truly interoperable database system and a remote DBMS interface that accesses multiple databases one at a time. The latter type of system, for instance, would not allow the join operation across two databases whereas an interoperable system would allow such an operation [2]. Most of the current production-level systems access multiple databases, one at a time.

The creation of an interoperable information retrieval system is a difficult problem as IR systems, whether local or remote, are generally both autonomous and heterogeneous. IR systems are autonomous in that each system is independent of other systems and is complete in and of itself, i.e, it does not need any other retrieval system to function. IR systems are heterogeneous with respect to data models, query languages, and schema. In addition, an IR system manages a set of databases, the members of which may be autonomous and heterogeneous (see Section 3).

3. Axis for the Characterization of Information Retrieval Systems

The two axis indicate the ability of users to impose views in a network environment of simultaneous access to heterogeneous databases. Networks have made available not just a tremendous number of databases, but a multitude of different types of databases. Users now have easy access to bibliographic databases, electronic bulletin boards, newsgroups, electronic mail, full text databases, etc. Thus, we refer to a data set as the set of databases that a user might wish to access, and this data set may consist of multiple types of databases.

As indicated in Section 2, above, information retrieval databases tend to be autonomous and heterogeneous. An autonomous database is one that is complete in and of itself, such that updates to other databases do not impact on its integrity and updates to this database do not impact other databases. For the purposes of this characterization, we define two types of database heterogeneity, semantic and content.

Databases are semantically heterogeneous when there is a disagreement about the meaning, interpretation, or intended use of the same or related data [2]. There are many different types of semantic heterogeneity. For example, two bibliographic databases are semantically heterogeneous if one names an attribute "keyword" and the other names the same attribute "descriptor". They are heterogeneous if the values for the "size" attribute of an item in one database is in centimeters and in the other database it is in inches. They are heterogeneous if one database has a single category for both conference proceedings and books and the other database has separate categories for proceedings and books. In this last instance, the GET command on ORBIT, which provides the frequency of single or multifield values for a retrieved set, would return inconsistent results if used to determine the number of retrieved items of type book from these two databases.

Databases are content-heterogeneous if predicates that describe the meaning of the data stored in the database differ substantially from one database to another. For example, databases of employee records, bibliographic records, and electronic newsgroup records are content heterogeneous. However,

information needs might require access to all three databases, simultaneously, and it is reasonable to expect such access.

Many of the available databases are autonomous (independent of each other) and heterogeneous. As such, there is no global view of these databases as data sets. Although there is some transparency at the query level (see Section 3.1), there is less evidence of transparency in presenting the results to the user. Therefore, the user has to interpret and integrate the results of each query as best they can.

3.1 The Query Axis

This axis reflects the ability of users to impose views at the time of query. This is really a function of how tightly the retrieval command language is coupled to the database structure. This could range from having a single command set for a single database to having a single command set for access to multiple heterogeneous databases.

A single command set can be used for access to a data set consisting of a single database or a homogeneous distributed database. This assumes homogeneity at both the target retrieval system level and at the database level. At the system level, this implies the same data model and the same query language. Thus, the command language can be tightly coupled to the database structure and the user can impose a view by structuring a query appropriately.

A single command set can also be used for access to heterogenous databases as in the Euronet-DIANE network [8] in which each system uses the Common Command Language. Although each system uses the same query language, they do not necessarily have the same data model and, as a result, not all systems will have the same functionality. Thus, it would be somewhat more difficult to impose a view through the query as the target system may not have the appropriate functionality.

A switching language, although it presents the user with a single command language, translates the request into the language of each target system. The user can only access those systems which the switching language supports. Typically, the target systems have data and command similarity. Unfortunately, not all target languages have the same functionality and not all target databases have the same access points. This may result in an inconsistency of results. As such, it is quite difficult to impose a user view through the command set.

3.2 The Presentation Axis

This axis reflects the ability of users to have the retrieved information presented in such a way as to best meet their individual requirements. This can be measured in increasing complexity from a single view defined by the database server, through multiple views of the same data but within the same model, to multiple views in different models.

In an interoperable retrieval system environment, there is no global view. Therefore, the retrieved information is presented through the individual views of each database, through a mediator [9] which stands between the user and the databases to fuse various views together, or through an imposed view of the user's design.

In some systems, it is not difficult for a user to impose various views on a database, if the views and the database are within the same model. For instance, in the relational model the user can develop views by selecting different sets of attributes to be part of the view. Each such view is consistent with the underlying model.

However, the user can also impose different views of a database, based in different models [10]. For instance, a bibliographic database can be viewed as a hierarchy of parts, as a serial data stream, or as a table of attribute values extracted from either the database or from retrieved items and

placed in relations. This gives the user a great deal of flexibility, but perhaps at the cost of increased complexity.

In addition, the ability to impose views depends on whether the view can be imposed dynamically on the remote database so that queries reflect the view or only on items after they have been retrieved and cached locally. Note that this is not simply downloading records for future use; it implies retrieving records and imposing views on those records by the retrieval system as part of the query process and search process. Locally cached records may or may not be available for future use.

4. Characterization of Various Systems

In this section, various information retrieval systems are characterized using the above attributes. The systems have been divided arbitrarily into early systems, current production systems, and new approaches. In no way is this meant to be a review of the many systems that are available. These systems have been selected simply for the purpose of illustrating characterization by these attributes.

4.1 Early Systems

Although users have been able to dial in to online retrieval systems for the past 20 years, it is really only in the past 10 years that systems have appeared in which computers communicated with computers in a network environment for information retrieval.

4.1.1 MESSIDOR

The MESSIDOR system [11] was probably the first system to allow a user to work in a single query language with several bibliographic databases, each using a different query language. This switching language is based on an early draft of the Common Command Language. The query is translated into the languages of the target systems and broadcast to all of the target systems simultaneously. The user may also search a single database in the native command language of the target system. Intermediate results consisting of the numbers of documents found in each database are integrated into a single display. Display of retrieved documents allows some view capability through field selection. The data set is semantically heterogeneous.

This system could be characterized on the query axis as either using a switching language or permitting access using the native language of the target system. On the presentation axis this system uses a mediator for intermediate results and either the switching language or the native command language to impose a presentation view.

4.1.1 PSI and CONIT

The PSI [12] and CONIT systems [13] were similar in that they provided a common interface to multiple bibliographic databases. They are simpler than the MESSIDOR system in that they accessed only one system and database at a time and did not try to interpret the results. PSI is a microcomputer based system that accessed any database on either DIALOG or the Canadian system, CAN/OLE. Early versions of CONIT were mainframe based and accessed DIALOG, ORBIT, and two implementations of Medline, one at NLM and one at SUNY at Albany.

Both of these systems could both be characterized on the query access as using a switching language and using the presentation view supplied by the remote system.

4.2 Current Production Systems

Current production systems place a great emphasis on online searching aids [14], including front ends, gateways, intelligent intermediaries, post processing, etc. This paper does not deal with intelligent intermediary systems and only looks at a few representative systems. We also assume that the new databases formed from vertical slices of other databases in order to serve specific markets can be treated as just more databases.

4.2.1 The Intelligent Gateway

The Intelligent Gateway [15] has been under development since 1975 at the Lawrence Livermore National Laboratory. This system provides the user three different ways of querying a target system; in the target system's native mode, through a switching language which provides a common command language to the target system, and through a fully automated search and retrieval procedure for routine tasks. Simultaneous connection to various systems allows the user to move from one system to another as needed, but each connection is kept separate. This does allow, however, a user to interrupt a database search to retrieve information from another source and then to resume the database search. Thus, the data set is content heterogeneous. Post processing tools include reformating to a common format to permit merging of results from different sources.

This system could be characterized on the query axis as either using a switching language or permitting access using the native language of the target system. On the presentation axis this system uses a mediator for merging results from different databases or the native command language to impose a presentation view.

4.2.2 Euronet-DIANE

Euronet-DIANE [8] provides access to multiple databases at various sites in Europe. A central server called Echo provides users with information about the network and its databases. All servers provide access though the Common Command Language.

A system operating in this environment could be characterized on the query axis as either using a switching language or permitting access using the native language of the target system. On the presentation axis this system either accepts the view presented by the target system or uses the native command language to impose a presentation view.

4.2.3 EasyNet

The EasyNet gateway system [16] provides access to over 850 databases at various sites. Its latest version provides a common command language to many of these databases. Its Scan option permits simultaneous searching of groups of databases. The databases are grouped by subject and the user selects, via a set of menus, the most appropriate subject group. The query is run against each database in the group and the postings are brought back and displayed to the user. The user then selects one database at a time to view the actual results.

For access to multiple databases, this system could be characterized on the query axis as using a switching language and on the presentation axis it provides the view presented by the target system.

4.3 New Approaches

The one thing that all of the following systems have in common is a move towards developing interoperable databases, i.e, access to multiple databases acting as though they were one virtual database.

4.3.1 Wide Area Information Server

The Wide Area Information Server (WAIS) [17] is an architecture for access to content heterogeneous databases in a network environment. In response to queries, documents are retrieved and cached locally in dynamic folders. Dynamic folders are sets of documents associated with a query. These folders can be updated with new documents either actively by the user requesting it or passively by the query associated with the folder being automatically executed on a periodic basis. These queries can be broadcast to multiple databases which have been identified through a directory of databases. As WAIS is based on the Z39.50 protocol, a client can also act as a server by allowing its dynamic folder(s) to be accessed by other clients. Many different types of databases can be accessed using this protocol and WAIS does not specify the query language or the format of the retrieved records. However, as the returned records are cached locally, there is flexibility in processing these documents and in imposing user views on them. WAIS can also be viewed as a large-scale hypertext system by allowing links to be established at runtime and across many databases and systems.

Systems based on this WAIS architecture can be characterized on the query axis as either providing a switching language or using the native language of the target system. On the presentation axis, user views can be imposed on local cache but not on servers.

4.3.2 A Distributed Indexing System

This system [18] is based on the idea of brokers. A primary site broker controls access and updates relating to a primary bibliographic database. An index broker indexes specific primary databases at multiple sites. The indexes are generated by a generator query which is registered at each primary site. The topics brokers group indexes and primary databases on related topics.

A user query is translated into a common query language and the topic broker database is accessed to find appropriate index brokers that point to databases of interest. It is hypothesized that such a system can be used to support dynamic instantiation of nodes and links in a hypertext system.

This system can be characterized on the query axis as using a switching language. On the presentation axis, the view is supplied by the generator query that establishes the index broker.

4.3.3 Daltext

Daltext is a prototype system developed at Dalhousie University to support research into text-based retrieval systems. It is based on transient-hypergraph model for data access [19]. Although not currently functioning in a network mode, it supports access to content heterogeneous databases. Its hypertext interface supports access to multiple databases of content heterogeneous natures, and user-imposed views on these databases [11]. Database items are retrieved and cached as sets which can then be manipulated as required. Of special note with this system is that it allows multiple user-imposed views of the data. The user can view a database as a hierarchy of parts, as a serial data stream, or once items have been retrieved data can be extracted and the items can be viewed as a table of attribute values.

Figure 1 shows user imposed views in different data models of a single bibliographic database consisting of references to articles appearing in the *Proceedings of the 1982 Conference on Computer-Human Interaction.*[2] The **sets of items** window indicates that one set has been retrieved based on a string search of the database for all items containing the string, "Interactive". The **set1** window lists

[2]Data from The HCI Bibliography Project, The Ohio State University, Columbus, Ohio

the items in the retrieved set. The **CHI82** window displays items retrieved from the CHI82 database.

In Daltext, the user is permitted to extract data based on one or more user-defined attributes, and place the data into a table which is a universal relation (UR) [20]. The definition and instantiation of the universal relation are dynamic. Multiple attributes may be defined for either the target database or a retrieved set of items and the universal relation may be modified by adding new attributes.

In Figure 1, each tuple (topic, title, author) shown in the window **UR TABLE** consists of data extracted from an item in the database while each tuple in window **set1 UR table** consists of data extracted from a retrieved item in **set1.** All of the browse and query operators available within Daltext are applicable within this relational view of the data. Each entry in these tables is associated with a corresponding database item, which can be selected and displayed in the **CHI82** window. These relations may be stored in an underlying relational DBMS to provide a persistent data view.

In addition to the relational view of the data, the user can impose a view in which both the database and the extracted data are viewed as an hierarchy of parts. A grammar can be used by the user to define such an hierarchy and also to query the database and browse through the extracted data (21). The hierarchical view can be instantiated dynamically at the direction of the user and need not reflect the structure of the original database. A query based on the hierarchy defines a new set of nodes within the transient hypergraph. The user can build up as much or as little of the hierarchy as is needed for a session and the hierarchical view, once defined, can be used both for accessing data from the database and for presenting and browsing the data within the context of the sets of the the transient hypergraph. The **UR objects** window presents a graphical view of the universal relation items.

Figure 2 shows better the use of the hierarchical view of a database. Using such a view of the data, the user has created two sets of nodes for browsing, as shown in the **sets of items** window, where a node is defined relative to the hierarchy. The first set contains nodes, where a node is defined as the subtree "book", that are instances of class c200. The second set, shown in the window **set2,** contains nodes of book parts that are instances of class c100 in which the word "Human" occurs in the title part. Browsing from **set2** the user has selected the first node for display and this is shown in the window **object-view**.

Users can impose views on content heterogeneous databases as well. Even if the databases are content heterogenous, the semantics of a view will be valid across multiple databases, even if the implementations of those views differ. The retrieved items are integrated via the transient-hypergraph model.

This system can be characterized on the query axis as permitting user imposed views on the databases. On the presentation axis, the system permits multiple views to be imposed and these views may be based on different models.

5. Conclusions

The query and presentation axis have been presented as a possible framework for viewing the field of information retrieval systems in a network environment. A framework is helpful for viewing the state of the art, serving as a reference point for future and unforseen developments, and highlighting areas where creative thought and development are needed.

In characterizing the handful of systems discussed in this paper, it is clear that there are a number of different approaches being taken to deal with the problems of access and presentation of multiple heterogeneous databases in a network environment. Typical approaches to querying seem to

rely heavily on the use of switching languages which are difficult for users to structure effective views in without knowing the functionality of the target systems. These switching languages also seem to be confined to switching within within the same model. Presentation tends to rely heavily on simply presenting the view from the target system, perhaps with selected fields suppressed (or presented). There is more flexibility in presentation if retrieved data items are downloaded first and the views imposed locally.

Within this framework, it is obvious that users, or their client machines, have to have descriptions of the views of the various databases that they wish to access. Perhaps this can best be done through the use of a common language such as in the Euronet-DIANE system with self-describing databases. In this instance, the instantiation of a query view would not be done at the client end and passed to the server, rather it would be done at the server end and have access to the description of the database.

On the other hand, the user should be able to impose a view in the model of their choice. This is difficult to reconcile with the concept of a common language. The Daltext system does allow the user to impose views in different models. However, in order to do this the user must have some idea of what the raw data items look like in the target database.

Thus, an area of interest might be the coupling of various views of the database(s) with the concept of self-describing databases.

REFERENCES

[1] Flynn, M. 1966. Very high-speed computing systems. *IEEE Proceedings*, Vol. 54, 1901-1909.

[2] Sheth, Amit P. and James A. Larson. 1990. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys*, Vol. 22(3), 183-236.

[3] King, Tim. 1991. Critical issues for providers of network accessible information. *EDUCOM Review*, Vol. 26(2), 29-33.

[4] Lynch, Clifford A. 1990. Information retrieval as a network application. *Library Hi Tech*, Vol. 32(4), 57-72.

[5] National Information Stantdards Organization. 1988. *American National Standard Z39.50, Information Retrieval Service Definition and Protocol Specifications for Library Applications.* New Brunswick, NJ: Transaction Publishers.

[6] International Organization for Stantdardization. 1990. *Documentation -- Search and Retrieve Protocol Specification. ISO/TC45/SC4/WG4. ISO/DIS 10163.* Vienna, VA: Omnicom.

[7] Litwin, W., Mark, L. and Roussopoulos, N. 1990. Interoperability of multiple autonomous databases. *ACM Computing Surveys*, Vol. 22(3), 268-293.

[8] Hensley, Colin. 1989. Eupopean Community Databases: Online to Europe. *Database*, Vol. 12(6), 45-52;

[9] Silberschatz, A., Stonebraker, M. and Ullman, J. 1991. Database systems: Achievements and opportunities. *Communications of the ACM*, Vol. 34(10), 110-120.

[10] Watters, C. and M.A. Shepherd. 1991. *Shifting the information paradigm from data-centered to user-centered.* Tech. Rep. No. CS90-03. Halifax, Canada: Dalhousie University, Dept. of Mathematics, Statistics & Computing Science.

[11] Moulinoux, C., Faure, J.C. and W. Litwin. 1983. MESSIDOR: A distributed information retrieval system. In *Lecture Notes in Computer Science, Vol. 146: Research and Development in Information Retrieval,* Edited by Gerard Salton and Hans-Jochen Schneider. pp. 51-61. New York: Springer-Verlag.

[12] Shepherd, M.A. and C. Watters. 1984. PSI: A portable self-contained intermediary for access to bibliographic database systems. *Online Review*, Vol. 8(5), 451-463.

[13] Marcus, R.S. and J.F. Reintjes. 1981. A translating computer interface for end-user operation of heterogeneous retrieval systems. I. Design. *Journal of the American Society for Information Retrieval,* Vol. 32(4), 287-303.

[14] Efthimiadis, Efthimis N. 1990. Online searching aids: A review of front ends, gateways and other interfaces. *Journal of Documentation*, Vol. 46(3), 218-262.

[15] Burton, Hilary D. 1989. The Livermore Intelligent Gateway: An integrated information processing environment. *Information Processing & Management*, Vol. 25(5), 509-514.

[16] Meyer, Daniel E. and Den Ruiz. 1990. End user selection of databases -- part IV: People/news/general reference (on EasyNet). *Database*, Vol. 13(October), 65-67.

[17] Kahle, Brewster. 1991. *Wide Area Information Server Concepts, Version 4, Draft.* Available by anonymous ftp from: think.com:/wais/wais-8-b1.tar.Z

[18] Danzig, P.B., Ahn, J., Noll, J. and Katia Obraczka. 1991. Distributed indexing: A scalable mechanism for distributed information retrieval. *Proceedings of teh Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval.* pp. 220-229.

[19] Watters, C. and M.A. Shepherd. 1990. A transient hypergraph-based model for data access. *ACM Transactions on Information Systems*, Vol. 8, 77-102.

[20] Hartzman, C.S. and C. Watters. 1990. A relatonal approach to querying data streams. *IEEE Trans. on Knowledge and Data Engineering*, Vol. 2, 401-409.

[21] Salminen, A. and C.R. Watters. 1991. *Grammar based access to hypertext.* Tech. Rep. No. CS90-02. Halifax, Canada: Dalhousie University, Dept. of Mathematics, Statistics & Computing Science.

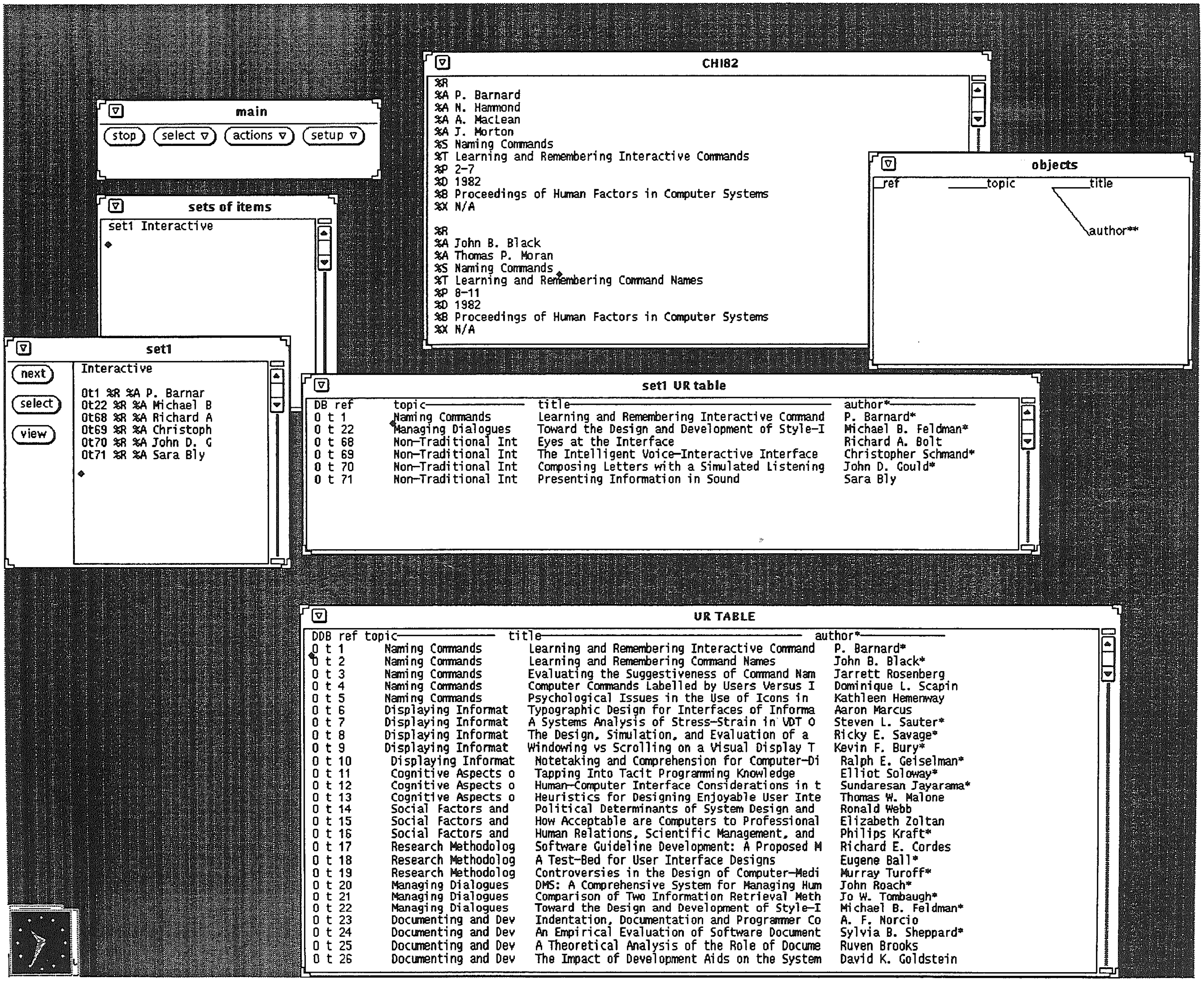


Figure 1. Relational View of Data from Daltext System.

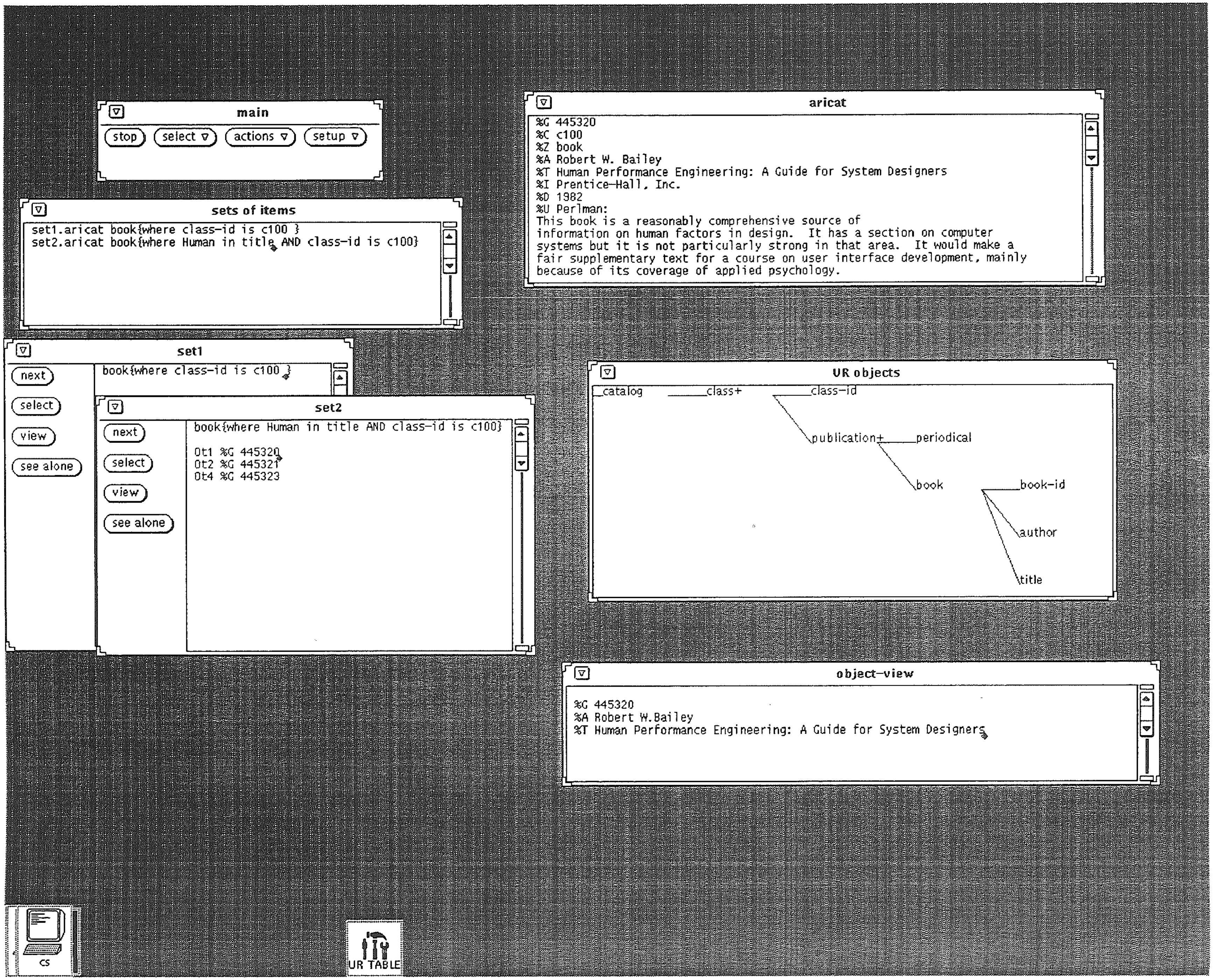


Figure 2. Hierarchical View of Data in the Daltext System.

"From Security to Serendipity, Or, How We May Have to Learn to Stop Worrying and Love Chaos"

Joseph W. Janes & Louis B. Rosenfeld
School of Information & Library Studies
University of Michigan
Ann Arbor MI 48109

INTRODUCTION

Our first information storage and transmission system was ourselves. What information there was (where the buffalo herd was, which berries would kill you and which ones wouldn't, how to build a fire) was conveyed through gesture, example, and eventually language, and was stored in our memories.

In those halcyon days, retrieval of information was relatively easy. You either remembered it or somehow made your wishes known, and if anybody you could find knew what you wanted, you got your answer.

It's been downhill ever since.

The development of representation schemes (drawing, writing, numerals) started us down the long road to where we are today--engulfed by an inconceivable and often undifferentiated mass of stuff in which it can be virtually impossible to find what you're looking for. Yet our Paleolithic desire to get and know everything persists.

We hope to show why, in our current state, it is becoming more and more difficult to get all the available information on a given topic, how chaotic the process and environment have become, and the implications of these developments.

TECHNOLOGY HIDES INFORMATION

When all information was internal, it was all immediately accessible directly from memory, and you could be almost certain of retrieving what you wanted. In modern information retrieval terms, recall = precision = 1. In preliterate or nonliterate societies, storytellers, troubadours and historians developed incredible techniques of memorization and retrieval, often involving rhythm, song, and imagery. If you have no way to represent your information, you have to remember it all.

Information technologies have created two related but quite distinct problems of information retrieval: <u>scale</u> and <u>distribution</u>. These technologies have enabled the production and storage of ever greater amounts of information and ever greater distribution of information in distinct locations. Each of these has served to make more and more of this information "invisible" to a searcher--less and less like when it was all in our brains and directly accessible. It is crucial to note that scale and distribution are two separate issues--they are often mistakenly elided, and doing so misses an important aspect of the problems they generate and may prevent the development of potential solutions.

When writing was developed, memory techniques were no longer necessary. It was then possible to store information more permanently, with less possibility of distortion or loss, and you didn't have to be within shouting distance of someone to know what they knew. At the same time, though, it was also less accessible: once a significant collection of written texts was compiled, if you wanted information contained in one of them, you had to find the right one--not always a trivial task. It was no longer possible to access the information directly, and that mass of texts also, as a by-product, "hid" some information--a text could exist which no one knew about which contained information someone needed, but it would be very unlikely that it could be found. Recall and precision began to dip. Still, though, if there were not many texts, one could conceivably know or know of them all.

Then, in the fifteenth century, printing using moveable type became widespread in Europe. In a very short time, the number of texts and number of different texts exploded. The pile became much larger, and it again became harder to know what was available and what information was contained in which physical vessel.[1] This was not a qualitative difference from the previous situation, but rather a massive change in scale.

A change in distribution occurred as well, however. With more books available, more collections--more piles--arose. It was no longer simply a matter of finding the right book, it became a matter of finding where the right book might be located.

The next change came with technologies which permitted alternative methods of representation of textual information--microform but especially digital formats. These formats not only allow piles to get bigger quicker, they also permit a virtual pile. Absent some method of organization and access, information becomes merely series of electrical pulses or magnetic spots.

Now the information has receded from our grasp a further step. The days of "knowing all there is to know" seem very far away--indeed, the fears of information overload, infoglut and information anxiety are commonly expressed. At the same time, we are afraid of getting too much information and not getting enough--or at least not getting or being able to get all we wanted. Digital information is inherently invisible without the aid of technologies, and searching for information stored digitally (in, say, online catalogs, CD-ROMs or online retrieval systems) often results in a satisfactory retrieval set. Often, though, that retrieval is accompanied by a vague unease on the part of the searcher that there's more to be had in another database, or on a related topic, but you just couldn't find it. Whether end-users experience this same unease is an open, and interesting, question.

Add to this the notion that a document or a text may never be "finished". Digital formats permit dynamic documents which may continually change and never reach a final or printed form. The problems of retrieving such information are clear: the document which contained the information you seek may be there and you may find it, but the information may have been removed or edited since the last time you saw it, or since it was referred to you.

The most recent step down this path is the increasing storage of information in highly distributed ways using wide-area networks. The analogy is clear: the Matrix is to

[1] Changes similar to these were produced with the invention of devices capable of recording sound, images, and moving images.

DIALOG and MELVYL as printing was to writing--a change in scale but primarily the enabling of vaster duplication, repackaging, and distribution of information. The virtual pile just got broader and more fragmented.

As each of these information technologies (writing to printing to electronic to network) has been introduced, more and more information stored has become invisible; it not only is harder and harder to retrieve what you want, it becomes, in practice, harder to know what you haven't got, because you can't "see" (i.e., access) it. None of which, though, stops us from wanting to get all we want and no more.

CHAOS PRODUCES STRUCTURE

Reading the above account would make one wonder how we ever came out of the caves--how could we find anything? Of course, we have developed a series of methods and schemes to deal with the invisibility problems, motivated by scale and distribution issues, over the centuries.

As writing became popular, so did the idea of titles and authorial credit to identify which scrolls, clay tables or papyri were which and, perhaps, which ones were most likely to have the information you wanted. Early libraries employed organizational schemes of increasing complexity and sophistication[2]. The library at Assurbanipal (1668-626 B.C.) had a crude shelf list with finding aids such as title or opening words and location symbols, and the Greeks began using author's names as identification, although items were organized chronologically or by accession order. During the medieval period, these lists slowly evolved, organizing works by broad subject, but still resembling shelf lists more than catalogs. The earliest attempt at a union catalog was in England in the thirteenth century, and author indexes began in the fourteenth century.

Widespread mechanical printing, begun in the fifteenth century, led to what was probably the first general catalog designed to be used as a finding list: the Bodleian catalog of 1620. It was arranged by author and short title (for anonymous works) in dictionary format. This work greatly influenced cataloging as standards tentatively began to arise in the seventeenth and eighteenth centuries.

Cataloging, we see, has been developed to deal primarily with the problem of scale. The solutions which have emerged to deal with the problem of *distribution* fall into two categories: buy everything you can get (the enormous central library idea), or get access to everything you can get (interlibrary loan systems). Large rich libraries could afford to employ both of these strategies; smaller and poorer institutions fall back on the second. In the print world, the problem with distribution of information is that sometimes you didn't have enough access to what you wanted.

The use of computers to store and retrieve information has led to a number of kinds of data structures (data types, arrays, stacks, deques, queues, strings, linked structures, hyperlinks, files) as well as a variety of novel techniques for searching (Boolean searching, full-text retrieval and keyword searching, expert systems,

---

[2]The following brief discussion of the history of cataloging takes most of its content from Eugene R. Hanson and Jay E. Daily's excellent article "Catalogs and Cataloging", in Volume 4 of the *Encyclopedia of Library and Information Science* (New York: Marcel Dekker, 1970), p. 242-305.

probabilistic and statistical techniques, hypertextual searching, etc.). Largely, though, in organizing textual information, we have fallen back on notions borrowed from cataloging: subject headings, index terms, and a few extra bells and whistles.

As this curtain of invisibility has descended time and again, then, we have developed a variety of organizational structures to help us better to handle larger and more widely distributed information masses. And, of course, it's happened again.

The Cycle Repeats

The problem of scale, in its electronic incarnation, is known as information overload. In fact, for most of us, discussion of information overload has become almost passé. As we produce and consume documents in greater numbers, our electronic mailboxes overflow. However, there is a paradox: despite the flood of information that reaches us, we are overcome by the uneasy feeling that we might be missing something. The network has enabled us to solve the old problem of distribution, perhaps too well: the ever-widening distribution of network-based resources has begun to obscure the knowledge we seek. And, as in the past, we begin to search for some new structure to filter, catalog, index, tag, organize and otherwise make accessible the glut of information we've come to know as the Matrix.

Most of the efforts to bring order to the chaotic world of networked information can be described as attempts to "catalog the Internet". These efforts have concentrated on cataloging entire databases or collections of documents. This approach is reminiscent of what has traditionally been done in archives. And though many of these efforts hold promise, they may suffer the flaw of applying old tools, such as cataloging, indexing, and archiving, to new problems.

In the Library of Congress' MARBI Discussion Paper #54, "Providing Access to Online Information Resources"[3], a number of difficulties in cataloging networked resources are brought to light. When considering these resources, should "computer-mediated communication", such as electronic mail and bulletin boards, be cataloged alongside bibliographic databases? Some resources are as much tools or services as they are data sources; how should they be distinguished? Is an online information resource a document, a collection of documents (e.g., a Usenet newsgroup), or a collection of collections of documents (e.g., the entire Usenet)? And as authors often distribute as well as create their electronic documents, how are data producers and distributors to be distinguished?

While that paper raises several key questions, its suggested solution is modeled upon a MARC-compatible structured record designed to describe the networked resources. Such a highly structured record, however, is designed for a single, static medium (e.g., books), and as such is often quite inflexible. In many online resources, the subject content of a resource can change almost daily. In addition, the number of resources increases exponentially; the format and media of these resources are extremely heterogeneous; and the number of these electronic media are increasing as well. Will a structured record be sufficiently flexible to describe widely different resources, and is keeping track of widely distributed resources a realistic goal?

[3]electronic document prepared by the Library of Congress Network Development and MARC Standards Office for discussion at the 1992 ALA Midwinter Conference.

Similar approaches have been undertaken by OCLC and the Coalition for Networked Information (CNI). Like LC, CNI and OCLC are investigating the use of the MARC data file format, and as such are compiling lists of suggested data fields for electronic information to aid in the creation of catalog records. In addition, both organizations are considering developing and testing descriptive taxonomies for types of networked resources; CNI's Top Node for Online Resource Information[4] may employ existing subject heading schemes, such as Library of Congress Subject Headings. OCLC is undertaking a very similar enterprise.[5,6]

The creation of taxonomies suggests the categorization and indexing of these networked resources. However, the known difficulties of imposing a controlled vocabulary will be magnified in an electronic setting, as the rapid evolution of knowledge and growth in numbers of documents will likely outstrip the viability of most vocabularies. Additionally, the self-published nature of many electronic documents and collections, combined with the great costs associated with indexing, will make it difficult to coordinate and enforce, much less agree upon, any specific controlled vocabulary.

WAIS (Wide Area Information Servers) is a tool that employs its own structured record to describe the networked resources that it makes available. This approach, similar to the building of MARC records, is likely to encounter the same problem of inflexible record structures. Other approaches, such as the St. George and Barron lists of online catalogs and the *Internet Resource Guide*, have attempted to compile descriptions of networked resources, and have consequently suffered problems of currency and restrictive formats.

An alternative to the indexing and building of bibliographies, directories and meta-databases may be the creation of associations or navigational paths between collections, documents and parts of documents. World Wide Web is an effort to make possible hyperlinks between documents distributed over wide area networks. However, the explicit creation of these links is a manual task, similar to manual indexing in terms of time and labor costs, and may be less useful, as these links often reflect the associations of a single individual. Similarly, a navigational tool like Gopher relies upon associations, in the form of hierarchical categorization of resources, made by many individuals who differ in their views of how information should be organized.

These attempts to organize networked documents and resources are well-intentioned.[7] However, most are based upon the principles of archiving, cataloging, and indexing which have been used with printed information; as it becomes more difficult to find and differentiate documents, collections, and services in an electronic environment, these approaches may not be workable. When we call for bibliographic control of electronic information, we may be making the old mistake of fighting the last war.

---

[4]"Call for Statement of Interest and Experience" and "The Top Node for Online Resource Information: Editorial and Business Plan, 2nd Draft", both from the Coalition for Networked Information.

[5]"U.S. Department of Education Provides Grant for Internet Research", press release, October 2, 1991, OCLC Office of Research.

[6]personal communication, Erik Jul, OCLC Office of Research, February 25, 1992.

[7]We note in passing no earnest attempts to index network resources at the document level.

DISTRIBUTION DISCOURAGES US ALL

At present, there is no known method capable of describing and keeping track of the online resources of the Matrix, and none appears on the horizon.

We got our wish: in the print domain, distribution of documents reduced access and led to interlibrary loan and large collections. In the networked world, distribution of documents has led to near-total access to information items, and for the first time, we need to develop a structure to deal with massive distributed access. And until that structure emerges, the invisibility of widely distributed resources will continue to make us uneasy in our quest for the exhaustive search; recall could continue to drop to a point where it is an unachievable goal and an irrelevant concept.

Yet while the issue of distribution is paramount, the problems of scale remain. The consumer of information may never be aware of all relevant sources of information, but within the known sources he or she will encounter an exponential increase in the numbers of documents, which will result in information overload. And, assuming that recall will remain a primary goal of searching, the results of information retrieval will become less satisfying to the information consumer.

This is illustrated by the following example (see Figure 1): if we estimate that today's typical search retrieves 20% of all relevant documents, it might be fair to guess that the average search in the networked world of tomorrow might retrieve only 5% of all relevant documents. This lower number is due to both the increase in numbers of documents, and the invisibility of many new document collections to the searcher. However, 5% of an exponentially larger number of documents means the size of tomorrow's retrieval will dwarf today's 20% of all relevant documents.

	today	tomorrow
# total relevant documents	100	5,000
x recall	20%	5%
= # relevant documents retrieved	20	250

Figure 1.

If tomorrow's average information consumer experiences much lower recall, combined with large retrievals that exceed his or her futility point criterion (the number of retrieved documents the user is willing to look through)[8], "exhaustive" searching will become pointless. It is far more probable, however, that the consumer will never reach this point; decreases in precision due to the overall increase in amounts and variety of documents mean that the retrieval process likely will be abandoned long before 250 relevant documents are retrieved. If we don't come up with a mechanism to deal with the problem of information distribution in the networked world, the consumer will have to

[8]Blair, David C. *Language and Representation in Information Retrieval* (New York: Elsevier Science Publishers, 1990), p. 10-11.

learn to be satisfied with much lower recall and precision, or instead will be forced into adopting a radically different measure of successful information retrieval.

THE DEFAULT FUTURE

One potential model for future retrieval is serendipity. While security in retrieval means that the searcher hopes to achieve the highest possible recall, serendipity in retrieval will mean that the searcher would only hope to encounter *some* useful information. The information seeker would navigate, as well as search, through networked documents and collections, taking a path based on personal associations and tastes. As such, serendipity would be more a goal than a measure of successful retrieval.

Successful serendipity would not depend solely on the searcher's knowledge of accurate retrieval algorithms and operators. Productive navigation would also require knowledge of various new *approaches* to searching, some of which will be quite similar to ways in which we currently search, but some of which will be quite different indeed. Further, navigating will necessitate more flexibility and devoting a great deal of attention to these processes of searching, since the environment in which that searching will take place will be much more dynamic and distributed than those we are currently accustomed to.

For example, there are a wide variety of sources of information currently available on wide-area networks: sources we might term "traditional" (bibliographic databases, text databases, OPACs, etc.), tools (FTP, WAIS, software and software archives), guides on how to use the Matrix or even how to think about it (the *Internet Resource Guide*, *Zen and the Art of the Internet*, etc.), lists of sources and resources (the Barron and St. George lists of Internet-accessible library catalogs, lists of listservs), and the most interactive resource, humans (via email, Usenet, listservs). Each of these will be accessed and used in very different ways.

Moving from security to serendipity represents a major paradigm shift for users of information. The implications are far-reaching, especially in scholarly work, where the quest to know everything in a field (as evidenced by extensive literature reviews) is considered *de rigeur*. However, it should be noted that serendipity has already gained acceptance as a model for searching resources in one context, the Internet. In that context, it is referred to as "surfing".

Let us attempt briefly to discuss and characterize this potential new mode of searching. In so doing, we risk over-extending the surfing metaphor, but we believe it has some intuitive and descriptive power in this context. Searchers in this serendipitous "ocean" may be characterized by, among other things, their courage, their ability to swim, and the quality of their surfboard.

Some searchers will be timid, staying close to shore (known sources of information), clinging to their boards or even wading in the tide pools, seeking only information which is available there. Others will be hanging ten, going to unexplored areas, and searching widely. These people may well discover new sources, and in turn synthesize information in new ways. Most, we suspect, will take some middle ground, exploring both new and old areas, finding new things by luck more than by design, and searching primarily to satisfy their own needs, and not just for the sheer thrill of it.

The ability to deal with new and complex sources and great quantities of information might be likened to swimming ability: some searchers will be quite good at

working in such an environment and will stay afloat; others won't, and will drown. Finally, the equipment itself might have an impact on searching style and results: some of us have great boards and wax (hardware, software, and connectivity to facilitate searching), others won't, and will have to make do. Indeed, the water itself serves as a metaphor for the dynamic and varying quality of data we will encounter in the Matrix.

We don't believe that we're doomed to this fate. The concept of "surfing" as a searching style is an exciting one, but we must remember that it is merely a fallback strategy, which will only be necessary if mechanisms to organize and facilitate searching in the widely distributed world of the Matrix are not developed. Clearly, that organized Matrix would be the ideal, and could eventually lead us back to the beginning, to when all the information we wanted was available, visible, and accessible, and we could get at it immediately. Such a scheme would truly allow us to come full circle.

Until, of course, the next technology is developed, and changes it all again.

CONTRIBUTED PAPER PROPOSED TO THE 1992 ASIS MID-YEAR MEETING PROGRAM COMMITTEE

Author: Bor-sheng Tsai

Title: <u>Building Parallel Subject Knowledge Navigation Systems for Database Searching in Network Environments</u>

<u>Abstract</u>: Searching a very large scale database in a local area network environment, without proper knowledge navigator's support, can be a very time-consuming and frustrating experience. Building proper knowledge navigation systems and processing them in parallel with databases in local area networks certainly can improve and enhance searching. Aided by the electronic mail system, the information retrieval induces dynamic information processing. In this paper, the author illustrates ways for constructing such special knowledge navigation systems that can advise librarians and researchers during information resource searching, selecting, and receiving activities.

1. PURPOSE OF STUDY

1.1. Current Trends- NREN, Parallel Processing, and Multi-Tasking

Access to information resources through internet is a current national drive. [1] As we know, all networks have a tendency toward eventual chaos. In order to prevent the straying of networking activities toward chaos, precautionary measures must be taken . These may include the technical aspects and the political or administrative concerns, e.g. user-friendliness, intellectual freedom, information property right and access. [2] Some economic factors to be considered include network maintenance, costs of software and hardware, the extension of LAN to totally integrated office software platform for establishing electronic offices, etc. [3] One important observation, which is sometimes ignored by the academic world, is that the business world does not place "sharing information as their primary goal. Their goal is business." [4] In other words, group dynamic interpersonal computing, messaging, micromanaging, budgeting for short-term project, etc., are favorable because they are likely to survive overnight changes or downsizing required by many businesses, big or small. These situations somehow disturb and hinder the normal scholarly communication processes. Nonetheless, we generally agree that the network's intelligent interactive abilities would be able to ease the probable chaos through human-machine cooperation. Theoretically, an intelligent network must be simple, easy to use, quick, clear-cut, accurate, and versatile. But practically, it is only possible to

accomplish the Artificial Intelligence functionality guidelines which usually include: search heuristics, navigational aids, intelligent browsers, query languages supported through Expert System shells, very user-friendly visualization tools, and advanced graphics capabilities. [5] In other words, the forms of information that the system sends directly to the audience must be ready for immediate use and digestion. [6] This pursuit of putting a "virtual library" into scholars' hands with a working, affordable knowledge navigator is an important item for the nineties. To reach these objectives, a mainframe system and an intermediate program embodying custom human-machine cooperation is required. [7]

In the mean time, a pursuit of parallel implementation is also indispensable. Oddy and Balakrishnan has pointed out that "decisions about whether documents should be retrieved are not made in isolation, but on the basis of a holistic view of their positions in the densely connected structure of literature, terminology, and authors in a domain." [8] Consequently, large sets of data and graph structures are necessary.

In retrospect, current information retrieval processes using MARC-based OPAC (Online Public Access Catalog) or CD-ROM systems are mainly depending on Boolean and relational logic operations. The entry points are usually on author, title, subject heading, keyword, etc. These approaches are efficient but not always precise. Studies on user problems with OPACs suggested that an interactive system designed for the user's direct manipulation of objects on the display might improve the system's performance. [9] How to improve these retrieval processes for the purposes of easier access, and to enhance accuracy and completeness, are among the many challenging tasks faced by many information related professionals. These include librarians, information specialists, infopreneurs, information scientists, and information educators. One alternative toward the betterment of searching in a particular subject information field is the manipulation of a computer system's multi-tasking abilities. This multi-tasking operation involves coordination among DOS, Windows, electronic mail, and databases in stand-alone workstations or local area networks. Therefore, the building of information coordination systems is the focus of this article.

The author has developed a Machine Readable Mapping (MARM) metrics and a coordinating/shelling model. Both techniques are applying geometric coordination skills to control monitor screen representations under the boundary of a special subject knowledge field. The graphic resolutions achieved by both techniques allow system operators to navigate users in searching a very large scale database in a local/wide area network environment, such as LUIS (Library User Information System) of DALNET (Detroit Area Library Network). In the following sections, the author will describe: 1) the construction of a knowledge navigation system;

2) the topological configuration for coordinating strategic map-shells, 3) the transformation and evolution of knowledge maps; and, 4) the coordination of navigational systems in parallel operation with the DALNET and BITNET/INTERNET. Parallel operations with DOS and Miscrosoft Windows environments are also elaborated.

1.2. Definition: Knowledge Navigation System

Saracevic and Kantor recently stated that online searching is still an imprecise art. This is comprehensible especially when the boundary of a particular subject knowledge field can not be clearly defined. To improve searching, both investigators suggested that the information systems must depend "not on increased sophistication of technology, but on increased understanding of human involvement with information." [10] On the other hand, Howard Rheingold has indicated that immersion and navigation constitute the elements of a "personal simulator." [11] Thus, building a parallel subject knowledge navigation system can logically improve and enhance searching yet maintain integrities of all systems involved. More specifically, it is to be used as a parallel consultant for researchers during their searching of multiple databases in a local area network. The hypothetical model is set for building a subject knowledge navigator that is capable of converging a particular subject information spectrum through a controlled lens/filter. This results in a concentrated intelligent point which then diverges a particular spectrum through a second controlled lens/filter into proper distribution channels.

2. HOW TO BUILD A SUBJECT KNOWLEDGE NAVIGATION SYSTEM

2.1. Hypothetical Model

Basically, there are three types of interfaces: Graphical User Interface, Multimodal Interface, and Natural Language Interface. In terms of human-computer interactions, the Natural Language Interface is a preferable pursuit which aims at designing a language-based, dialogue-centered interactive system "possessing an artificial personality ... that simulates human conversation." [12] As James Geller pointed out, "multi-media interfaces with a graphics and a natural language component can be viewed as a Natural Language Graphics system without a host system." Inasmuch, the operation of the information coordination system matches what Geller's model of "an abstract graphics machine that receives the request to draw the object under the modality and at a location (x,y), then the function denoted by form is applied to the argument x and y. The location (x,y) will be the location of a privileged point of the object called the reference point." [13] Three types of expert systems are recognized: rule-based, semantic-net based, and frame-based systems. [14] The operation

described below consolidates these three systems.

The following are the methods adopted for building an automatic subject knowledge navigation system.

2.2. Constructing Navigation Maps

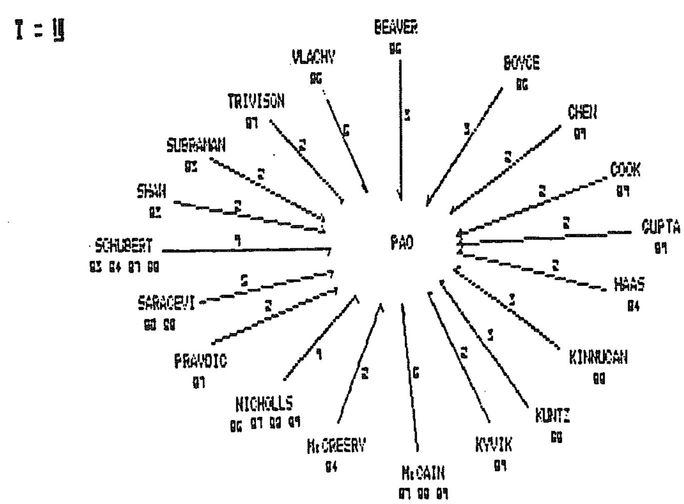
Usually a map, e.g. a tree diagram, can be drawn to guide data retrieval tasks. It has been found as more effective than those who receive other types of instruction. [15]

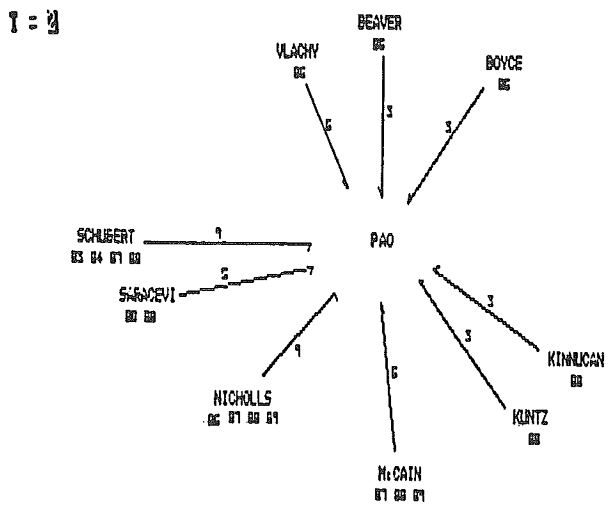
Three basic techniques are approached.

2.2.1. topological configuration

The establishment of basic configurations for a special subject knowledge navigation system depends solely upon the understanding of the OOPS (Object Oriented Programming System) concepts and techniques. Fundamentally, it requires thorough consideration and preparation on every basic element to be used by the system. A classification system for uniquely naming each object must be designed in the first place. [16] The produced general graphs may be dissected into several clusters according to chronological or topological order. To accomplish this, it requires periodical review and threshold control (denoted as T= number, where the number represents the frequency of citations). These graphs are used to guide and satisfy the users' needs during selectivities. The in-depth research levels can be concurrently determined by the user by raising or lowering the threshold. The basic model is illustrated by the two graphs below which demonstrate a top-down or bottom-up formation using threshold as a filter to focus on a particular citation frequency.

[Graph 1: Authors Who Cited Miranda L. Pao, 1980-1990]

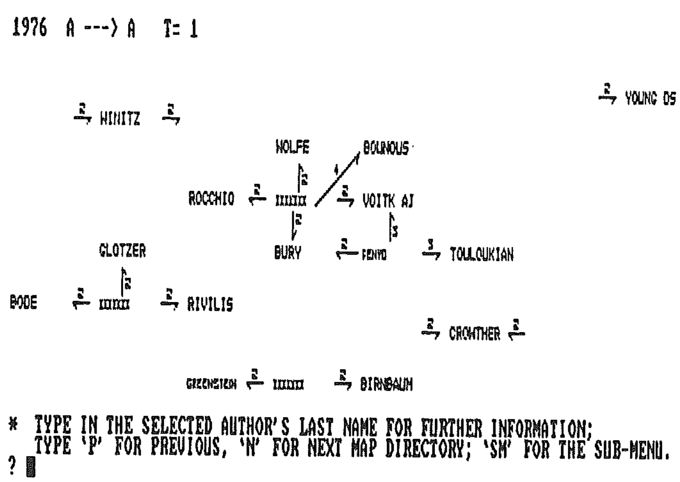


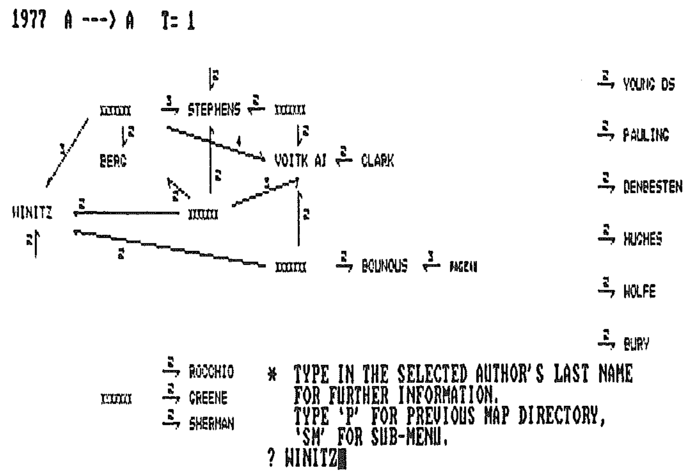


2.2.2. transformation and evolution

The transformation evolved in chronological order which can be observed and recorded through citation analysis that reveals long term scholarly communications relationships among the members in a particular subject knowledge field. The graphs shown below are some of the examples using citation analysis with the data taken from the SCI (Science Citation Index). Several significant figures retain intellectual leadership from one year to the next.

[Graph 2: Evolving Scholarly Communications]





2.2.3. parallel consultation

After preparing several sets of navigation maps, the next step is to annex them to the LUIS (Library User Information System). The connection from a PC to the LUIS System would allow the PC an added capability of toggling between the maps and programs with the LUIS as well as shelling back and forth among DOS, Windows, and other related software programs. Keyboard combinations such as ALT/ESC, ALT/ENTER function as toggle keys to move application programs from application to application (window or nonwindow-based), or from nonwindow-based to window-based setup. The following examples show the operation.

[Recall the navigation map needed. Choose a desired target. Check this target with LUIS]

```
 LUIS SEARCH REQUEST:  A=WINITZ
 BIBLIOGRAPHIC RECORD -- NO. 15 OF 15 ENTRIES FOUND

Greenstein, Jesse P. (Jesse Philip), 1902-1959.
  Chemistry of the amino acids <by> Jesse P. Greenstein and Milton Winitz.  New
York, Wiley <1961>
  3 v. (xxi, 2872 p.) illus., diagrs., tables. 24 cm.
  Includes bibliographies.
 SUBJECT HEADINGS (Library of Congress; use s= ):
     Amino acids.
 SUBJECT HEADINGS (Medical; use sm= ):
     Amino Acids

LOCATION: WSU SCIENCE/ENGG LIBRARY
CALL NUMBER:  547.75 G854c
```

[Connect from author search to descriptor/cited author search]

```
                    DESCRIPTOR : CITED AUTHORS

 EXCRETION:                                                        (E-6)

ATTEBERY(A2); BAIRD(B1); BERG(B4); CROWTHER(C7); DEITEL(D3); FELIG(F2); GLOTZER
(G3); GREENE(G6); HABTE(H5); KOUMANS(K5); MOSS(M4); SHERMAN(S3); SMITH(S6);
WEINBERGER(W2); WINITZ(W4); YOUNG(Y1).

*   TYPE IN THE  CODE   AFTER EACH AUTHOR'S NAME FOR FURTHER INFORMATION.
    TYPE 'M' TO RETURN TO THE MAIN MENU; 'SM' FOR THE DESCRIPTOR SUB-MENU.
?
```

This operation would greatly enhance the abilities of a PC to support the user with an extra guideline which can be recalled in parallel with the online database currently in use. This supportive guideline does not interfere with the current online operation but serves as an aid or a mirror which helps the user with various reflections that are derived from the preformed subject knowledge field. In this case, these reflections can include navigation maps, index links, automatic instruction programs, database management systems, hypertext-based software, and other related applications software.

[Using the topological map-shells based upon Pao's citations, exemplified in Section 2.2.1., a similar approach could result in obtaining the following information from LUIS]

```
 LUIS SEARCH REQUEST:  A=PAO M
 BIBLIOGRAPHIC RECORD -- NO. 4 OF 4 ENTRIES FOUND

Pao, Miranda Lee.
  Concepts of information retrieval / Miranda Lee Pao. -- Englewood, Colo. :
Libraries Unlimited, 1989.
  xvi, 285 p. : ill. ; 24 cm.
  Bibliography: p. 253-269.
  Includes index.
 SUBJECT HEADINGS (Library of Congress; use s= ):
     Information retrieval.
     Information technology.
     Library science--Technological innovations.

LOCATION: WSU PURDY/KRESGE LIBRARY RESERVES  (Circulation is restricted)
CALL NUMBER:  Z 699 .P29 1989
    Not charged out. If not on shelf, ask at Circulation Desk.

            FOR ANOTHER COPY AT THIS OR ANOTHER LOCATION, press ENTER

TYPE h FOR HELP, e FOR INTRO TO LUIS AND LOGOFF INSTRUCTIONS.
TYPE ANY COMMAND AND PRESS ENTER==>
```

```
 LUIS SEARCH REQUEST:  A=CHEN C
 BIBLIOGRAPHIC RECORD -- NO. 156 OF 220 ENTRIES FOUND

Chen, Ching-chih, 1937-
  Optical discs in libraries : use & trends / by Ching-chih Chen. -- Medford, NJ
: Learned Information, 1991.
  xv, 237 p. : ill. ; 28 cm.
  Includes bibliographical references and indexes.
 SUBJECT HEADINGS (Library of Congress; use s= ):
     Optical disks--Library applications.
     Optical storage devices--Library applications.
     Libraries--Automation.

LOCATION: WSU PURDY/KRESGE LIBRARY
CALL NUMBER:  Z 681.3 .O67 C46 1991
    Not charged out. If not on shelf, ask at Circulation Desk.


 LUIS SEARCH REQUEST:  A=NICHOLLS
 BIBLIOGRAPHIC RECORD -- NO. 156 OF 277 ENTRIES FOUND

Nicholls, Paul (Paul T.)
  CD-ROM collection builder's toolkit : the complete handbook of tools for
evaluating CD-ROMs / Paul Nicholls. -- Weston, CT : Pemberton Press, 1990.
  viii, 180 p. : ill. ; 23 cm.
  Includes bibliographical references.
  Includes index.
 SUBJECT HEADINGS (Library of Congress; use s= ):
     Data base selection--Handbooks, manuals, etc.
     Data bases--Evaluation--Handbooks, manuals, etc.
     CD-ROM--Evaluation--Handbooks, manuals, etc.

LOCATION: WSU PURDY/KRESGE LIBRARY
CALL NUMBER:  Z 699.22 .N5 1990
    Not charged out. If not on shelf, ask at Circulation Desk.


TYPE n FOR NEXT RECORD.  TYPE i FOR INDEX, g FOR GUIDE.
TYPE h FOR HELP, e FOR INTRO TO LUIS AND LOGOFF INSTRUCTIONS.
TYPE ANY COMMAND AND PRESS ENTER==>
```

2.3. DOS in Navigation

Learning to appreciate the functional and structural aspects in a computer, such as an operating system, data structures, ASCII codes, communication protocols, etc., are significant steps toward pedagogical instructing. [17] The DOS system is beneficial to the users for its capabilities in dynamic file creation, storage, retrieval, and transfer, as well as systems communications. To experienced users, DOS can also provide direct manipulations and links for file organization. In terms of navigation during information retrieval processes, DOS can provide similar operations as Windows, except DOS is more direct and logically clear which allows the users to have more control over the files interactively connected. Using DOS in file organization is more efficient and disorientation is less likely. Nevertheless, it takes skill and advanced training to fully understand and bring out the total capabilities of DOS. With the convenience provided by the Window-based software programs, to meet the demand of multimedia operations, the DOS practical usage for file organization, e.g. SHELL and EXIT, is seemly becoming neglected.

2.4. Windows in Navigation

The benefits that the MS-Windows offer are not the "mousy, graphical cosmetic of the interface" but its "dynamic data exchange (DDE)" capabilities for coordinating multi-tasks among application programs. [18] Using Window-based software programs allows users to quickly connect many applications that the programs provide. This enables the user to move back and forth among related software making the entire process more user friendly and altogether much more convenient and efficient. In terms of navigation during information retrieval processes, Windows can provide a quick switchboard-like operation that can toggle back and forth between software programs and online databases. This operation helps the users in accessing multiple useful advisory programs while continuing their online database searching. The keyboard combinations of ALT/ESC and ALT/ENTER play very significant roles in switching freely from application to application. One controlled factor is the coordination system which must be carefully designed by the system coordinator before it can be used by the searcher.

2.5. E-Mail in Navigation

The Electronic Reference Desk concept is gaining momentum. The e-mail bulletin system can be the integral part of the Electronic Reference Desk operations. [19] Borrowing the concept of virtuality, the reference librarian or information specialists can split or stretch time and space by way of telepresence and having conversation with distant callers. A new way of knowledge navigation is created. With the connection of the above other navigation systems, the whole information service chain is enhanced and complete. Starting from e-mail's call for information, through a coordination system's switch board that allow proper consultations with appropriate databases from the workstation or from the LAN-based OPAC, the relevant information can be coined out and delivered to the caller through an e-mail network. To accomplish this service, a new form of abstracting, indexing, and mapping skills for messaging has been developed. [20] It would allow the co-conversers to easily and clearly write/send and receive/read messages without involving ambiguities. During messaging, an "automated assistant" may be used to automatically respond to certain kinds of messages, and make suggestions. [21] [22]

3. HOW TO CONDUCT INFORMATION SERVICES

3.1. Parallel Interactivities among Navigation Systems-DALNET-BITNET/INTERNET

Parallel processing could be defined as an operation "under the philosophy of breaking a problem to be solved into manageable

tasks and allocating those tasks across several agents. ... Coordination is the glue that makes it possible for all the agents to work together." [23] It is similar to object-oriented programming in the sense that "programmers must be able to segment logically the task at hand to be able to run on multiple, simultaneous processors." [24] Following these guidelines, an experiment was conducted.

In general, the questions from the researchers may arrive in several ways: a) the researcher comes to the reference desk; b) the question is asked through telephone call; and c) the request arrives through e-mail. Upon receiving the requests, the reference librarian will decide to: a) lead the ordinary researcher to the bibliographic database and instruct the patron how to conduct simple searching; b) help the experienced researcher to locate the subject map which indicate the leaders in the subject field; c) contact the information coordination system [25] for identifying ready reference databases, such as LOTUS 1-2-3, dBASE, WordPerfect, PrintMaster, HyperTie, Strategic Mapping, Inf-B/N/F-Casting, Multi-Lingu/Cultur, IVD-Gilbert Files, IVD-Seizure Case, etc.; d) connect with the identified database for in-depth searching in order to retrieve condensed formulas which should cover the knowledge needed; e) direct strategies through maps for obtaining the needed facts and figures; f) toggle through the OPAC system, e.g. LUIS, with the identified database as supporting reference system, using SHELL/EXIT (in BASIC), ALT/ESC (in DOS, or for BITNET or DALNET) or ALT/ENTER (in Windows); g) decide to respond with answers or suggestions through e-mail or fax.

3.2. Information Coordination System Design and Operation

At the present moment, few librarians and information specialists see the needs and are not equipped with the system design and software programming techniques. It is difficult to conduct parallel knowledge navigation without proper training in these two regards. The best we can foresee is that librarians and information specialists will start learning, adopting, and operating the parallel supportive reference systems created by professional programmers, and be able to return feedback to the producers of these software packages.

4. RESULTS

4.1. Advantages

This experiment had identified several advantages. First, it helps users to advance themselves automatically to the specific areas they would not likely be able to reach when using a Boolean logic search. Secondly, the mirroring service could provide a highly condensed, topological digraph (directed graph) along with

several supportive hypertext-based or DBMS-based databanks to assist users during their searching. And, lastly, the information coordination system serves as a central controller which enhances the interfacing activities among the supporting databases, DOS and Windows, and online bibliographic databases such as LUIS or Wilsonline.

Some other advantages may deserve our attention. First, our subject knowledge navigator, programmed in a high level programming language such as BASIC, has received no virus attacks so far. Secondly, the system operation and information processing are in parallel, and consequently do not disturb each other. This prevent us from possible intrusion of the data structures and copyrights of the databases involved in our searching.

4.2. Limitations

Although the parallel knowledge navigation system of this type is simple and convenient to build, this experiment has encountered certain limitations. It provides indexing services to experienced users who are more familiar with the subject areas of which they are investigating. Therefore, the system might not be suitable for novice searchers. Furthermore, although experienced users only need a minimal training in recognizing search patterns before using digraph/hypertext/DBMS-based databanks and the information coordination system, they might not have patience to take extra steps. Regardless, since a total integrated multi-faceted system is still expensively underdeveloped, this parallel approach may be more cost effective. And, finally, the {IF-THEN-ELSE-[IF-THEN-ELSE-(IF-THEN-ELSE)]} bubbling or nesting might not cover all conditions required for an "intelligent" knowledge base. This might result in missing important links during information seeking processes. To compensate, the feedback from the interactive searching ought to be recorded and brought back to related supportive databases for necessary revisions and additions.

Some other services regarding multimedia may deserve our attention. For instance, some users may demand more powerful capabilities for file transfer, or cut-and paste links with other applications programs. [26] To satisfy this demand, the intellectual property and copyright issues need to be addressed first.

5. CONCLUSIONS AND SUGGESTIONS

Information technologies are constantly changing and improving. For the creativity and innovation to continue, one must not only keep up to date with the current technologies but also develop human-machine-system cooperative communications as well. This is

the area that the library and information education and professions must pay attention, in order to better our professional services, defend ourselves, and most importantly, build a strong scientific discipline. We have witnessed symptoms which have shown that the economy has affected the library and information science education. The consequences will unquestionably influence the information profession and the information society. It is clear that for the next few years, possible solutions and new directions ought to follow micro-computing and micro-management, while constructing global networks. To compromise and meet all compelling and oncoming challenges, this cost-effective approach applying parallelism as well as OOPS concepts and techniques seems desirable.

Further studies on CD-ROM databases linked with LAN are undergoing. The significant connection would allow users to access to multiple databases as well as permitting multiple users to search the same CDs or other network software packages through campus LAN, [27] or using a remote dial-up access to CD-ROM databases which are downloaded onto the online catalog. [28] Last but not least, the world of multimedia networking has yet much to be explored...

NOTES

[1] C. A. Lynch, "Internet Access to Information Resources," Annual Review of Information Science and Technology, 25(1990):263-312.

[2] B. Flanders, "NREN: The Big Issues Aren't Technical," American Libraries, June 1991, 572-574.

[3] R. Lockwood, "The Electronic Office," Personal Computing, May 25, 1990, 74-82.

[4] L. Flynn, "The Year of the LAN Arrived and It Was Nothing Like I Expected," InfoWorld, August 12, 1991, 41.

[5] D. N. Chorafas, and H. Steinmann, Intelligent Networks. (Boca Raton: CRC Press, Inc., 1990), 224.

[6] P. Scott, "The Organisational Impact of the New Media," Aslib Proceedings, 42(9): September 1990, 219-224.

[7] M. Mitchell, and L. M. Saunders, "The Virtual Library: An Agenda for the 1990s," Computers in Libraries, April 1991, 8-11.

[8] R. N. Oddy, and B. Balakrishnan, "Pthomas: An Adaptive Information Retrieval System on the Connection Machine," Information Processing and Management, 27(4): 317-335, 1991.

[9] D. Chiang, "Comparison of Direct Manipulation, Menu Selection, and Command Language as Interaction Styles for Online Public Access Catalogs," in Interfaces for Information Retrieval and Online Systems: The State of the Art, edited by M. Dillon, (NY: Greenwood Press, 1991), 7-19.

[10] T. Saracevic, and P. Kantor, "Online Searching: Still an Imprecise Art," Library Journal, October 1, 1991, 47-51.

[11] H. Rheingold, Virtual Reality. (NY: Summit Books, 1991), 100.

[12] M. Dillon, editor, Interfaces for Information Retrieval and Online Systems: The State of the Art, (NY: Greenwood Press, 1991), xv-xxvi.

[13] J. Geller, "Propositional Representation for Graphical Knowledge," International Journal of Man-Machine Studies, 34(1991): 97-131.

[14] R. A. Orchard, E. J. Reese, and M. R. Tausner, "On the Foundation of Knowledge Engineering," in Modelling and Simulation Methodology, edited by M. S. Elzas, T.I Oren, and B. P. Zeigler, (North-Holland: Elsevier Science Publishers B. V., 1989) 93- 126.

[15] J. M. Webb, and A. F. Kramer, " Maps or Analogies? A Comparison of Instructional Aids for Menu Navigation," Human Factors, 32(3), 1990, 251-266.

[16] B. Tsai, "Mapping Metrics for Subject-Object Coordinated Field Recognition and Representation." Paper presented at the Third International Conference on Informetrics, August 9-12, 1991, Bangalore, India.

[17] B. Sutton, "Overcoming Illusions: Thoughts on Learning to Use Small Computers," Online, July 1990, 86-89.

[18] A. N. Grosch, "Micros for the 1990's: An Update," Online, September 1991, 63-67.

[19] L. J. Kosmin, "Electronic Reference Desk: Multimedia Integration Opportunities," Proceedings of the 12th National Online Conference, edited by Martha Williams, 1991, 197-199.

[20] B. Tsai, "The Effectiveness Measurement of Electronic Mail Communications Within a Special Professional Community." Paper proposed to the 1992 ASIS Annual Meeting.

[21] L. B. Eliot, "AI & Telecommunications Make the Connection," AI Expert, March 1992, 8-10.

[22] M. Robinson, "Through a Lens Smartly," Byte, May 1991, 177-187.

[23] T. Nolen, "Parallel Processing for Problem Solving," AI Expert, February 1992, 35-40.

[24] J. Keyes, "Living in Parallel," AI Expert, February 1992, 42-47.

[25] B. Tsai, M. C. Chamness, and M. Y. Chau, "Managing Hypertext-Based Interactive Video Instruction Systems," in Studies in Multimedia, edited by S. Stone and M. Buckland, and published for the ASIS (NJ: Learned Information, Inc., 1992), Chapter 20.

[26] K. Scott, "E-Mail Enters the Client-Server Era," Data Communications, June 21, 1991, 20-21.

[27] C. Rhine, "Implementing a Local Area Network: The Effect on Present and Future Services at the Health Sciences Library," Information Technology and Libraries, March 1990, 102-107.

[28] K. S. Butcher, "The Rewards and Trials of Networking," Database, August 1990, 103-105.

Continuing Development of California State Packet Radio Project

Edwin Brownrigg
Executive Director
The Memex Research Institute

Abstract

This paper is about events that have not yet taken place. It relates to a set of plans that are rooted in several years of grant funded research in the California State Library Packet Radio Project. It describes the early goals of the project as well as the outcome that called for a fresh start, which is being undertaken in 1992 based in great part in new FCC rules under part 15.247 of Title 47 of the CFR. It describes radio modulation via spread spectrum as a key breakthrough for high-speed, wireless telecommunications for libraries. It names the principles,t heir relationships and their efforts to build demonstration networks in San Diego and San Francisco. Set forth are the project description, the plan of operation, the adequacy of resources and the evaluation plan.

Background

1992 will see dramatic evidence of the return-on-investment from research and development in the California State Library Packet Radio Project undertaken during the mid- and late-1980's. IBM, the Council on Library Resources, and principally the California State Library, all made grants to the University of California where Dr. Edwin Brownrigg was the principal investigator in a series of projects that explored the potential for packet radio — wireless, high-speed digital communications — among libraries.

While the early goals of the R&D were to demonstrate technological feasibility and to adapt extant FCC (Federal Communication Commission) rules to packet radio technology among libraries, the actual outcome was the need for a fresh approach. The project showed that the conventional radio technology and the standard digital encoding techniques of the time were becoming arcane approaches to achieving the R&D goals. In fact, the FCC was just then introducing into its rules (Title 47 of the Federal Code of Regulations) a new Part, 15.247, which allowed an exotic method of digitizing a radio wave, and which held promise for packet radio. The new FCC rules were a welcomed alternative to the politics of re-cycling Instructional Television spectrum for packet radio communication which were proving to be daunting.

Called spread spectrum, this new method of using radio to convey digital information under Part 15.247 presented several advantages for libraries and other civilian users, as well as large technological challenges to the telecommunications industry. The major advantages were that multiple users could share the same radio spectrum simultaneously, and that within prescribed transmitter power no user license would be required from the FCC. The challenges were to transfer spread spectrum technology from the military sector, where it had been perfected as a means of secure communication, into the FCC-regulated civilian sector, and, at a reasonable price.

R&D now under way involves a convergence of interest in California among The Memex Research Institute, Tetherless Access, Ltd. and special-interest user groups. Among the latter is the City of San Diego Public Library, which is using packet radio for a 1.54

The Council on Library Resources and Apple Computer, Inc. are the sponsors of the San Diego Packet Radio Project. One of the project's objectives is to prove that FCC Part 15.247 rules will work for libraries. Dr. Edwin Brownrigg of the Memex Research Institute, and Richard Goodram of San Diego State University are the co-principal investigators.

Tetherless Access, Ltd. and the Memex Research Institute are now seeking funding to deploy a network of some 600 packet radios in the San Francisco Bay Area for use by civilian groups, including libraries. This network will extend as far south as San Jose and north-east to Roseville. The network's radios will comply with FCC Part 15.247 as well as with an authorization from the FCC allowing Tetherless Access, Ltd. to apply FCC Part 97 rules (Amateur Radio) for the network backbone.

Together, the San Diego network and the Bay Area network are intended to demonstrate several technical features of packet radio: wireless wide area telecommunications; high data rates; last-mile access to the Internet; and, communication between such wireless networks through the Internet. They also are intended to demonstrate two precedent-setting public policy features of packet radio: common carrier by-pass for public benefit; and, use of the electromagnetic spectrum, a public good, in support of library service, also a public good.

Accordingly, the Memex Research Institute is seeking the voluntary participation of Bay Area libraries as nodes in the grant-supported wireless wide-area network. A single packet radio at a library will serve a local-area network within the library and gateway it to the wireless wide-area network extending to the Internet.

San Francisco Network Project Description

The goal of the overall project is to deploy in the San Francisco Bay Area a wireless, high speed, wide-area network comprised of 600 nodes of which libraries will account for 100. The means of achieving this goal will be packet radios designed and manufactured by Tetherless Access, Ltd., a California Corporation. A number of packet radio nodes in the proposed network will be gatewayed to the Internet at universities and research centers, thus being able to route data packets into and out of the Internet on behalf of the other nodes of the network. This will be the first network of its kind.

The other 500 nodes will be among other public civilian professionals, such as lawyers, and publishers. Some of the nodes will be mobile. Overall the Bay Area network will extend from San Jose to Roseville.

The project has the following main objectives:

1. To demonstrate the applicability of Title 47, "Telecommunication" of the Code of Federal Regulations, Parts 15.247 and 97 for packet radio operation, whereby the libraries will not be required to obtain a license from the Federal Communications Commission (FCC) to operate a packet radio transmitter/receiver.

2. To demonstrate that data rates of 1.54 megabits (T1) can be achieved wirelessly with "off-the-shelf" packet radios.

3. To demonstrate that no tariffed, common carrier circuits will be needed, and therefore that there is only a one- time cost of $3000 per node for the 100 library nodes.

4. To demonstrate that the computer to which a packet radio is cabled can act as a gateway from a library's local area network to the packet radio network and the Internet.

5. To demonstrate that every packet radio in the network is capable of dynamically routing and forwarding data packets on behalf of other packet radios.

6. To demonstrate how wireless nodes in the Bay Area network can exchange data packets with an extant five-node prototype sister network in San Diego by means of the Internet.

The Plan of Operation

The principal investigators propose to solicit among the libraries in the Bay Area 100 participants willing to install a packet radio with an Apple computer, and thereby to access the Internet. In addition 500 civilian sites will be established by means apart from this grant proposal. The solicitation will be a two step process. First, The Memex Research Institute will prepare a mass mailing to libraries in the Bay Area. The mailing will describe the project, ask for volunteers, and set a date and place for a meeting. At the meeting, functioning packet radios will be demonstrated and questions can be addressed.

Once the 100 libraries have been determined, priorities and installation schedules will be drawn up. The 100 libraries will place their orders through the Memex Research Institute. San Diego State University will be responsible for purchasing the computers and the packet radios. Like the Apple Computer, the Tetherless Access packet radio is designed to "plug and play." Therefore, a simple installation manual will accompany each packet radio, and each participating library will be responsible for installing it and interfacing it with the Apple computer. Both hardware components will be shipped directly from the manufacturer to the participating libraries.

The participating libraries will be responsible for the security and maintenance of the equipment that they purchase.

Adequacy of Resources

Only two physical resources obtain in this projects: a packet radio and an Apple Computer (model CI). The packet radios from Tetherless Access function under a trade-secret design that results in a 1.54 megabit spread spectrum encoding of a digital signal within a wide band of the electromagnetic spectrum. The packet radio interfaces with an Apple computer wherein trade secret software implements proprietary channel sharing protocol, TCP/IP (Transmission Control Protocol with Internet Protocol), and additionally manages the dynamic routing and forwarding of data packets in cooperation with neighboring packet radios.

Evaluation Plans

The co-principal investigators will study the performance of the packet radio network as well as the attitudes toward the network's performance and content among the librarians managing the network's respective nodes.

As to the performance of the network, special attention will be paid to how the network nodes react to and avoid congestion as well as the effective data rate versus aggregate data rate in a shared channel environment. In addition, the relative ease of installing and using the equipment will be reported. the packet radio netwocircuits. Also a financial model will be built that will compare the actual cost of the packet radio network against the theoretical cost of an equally performing tariffed common carrier network.

Because such use of the Internet will broaden its boundaries, it will also be desirable to survey the opinions of principals in the American Library Association, Coalition for Networked information, the Corporation for National Research Initiatives, the Internet Society, and the National Science Foundation.

The results of the evaluation will be published in a refereed journal.

Contacts

For more information, please contact Dr. Edwin Brownrigg, Memex Research Institute, 1220 Melody Lane, Roseville CA, 95678. Voice: 916.773.5910; Fax:916.786.7559; memex@calstate.bitnet.

NREN Economic Policy Issues

Thomas H. Martin
Syracuse University

Recently the High-Performance Computing Act of 1991 was signed into law establishing the National Research and Education Network.[1] By 1996 the NREN is to provide researchers and educators in academia, industry and government with appropriate access to supercomputers, computer databases, other research facilities, and libraries. Section 102 (c) (6) states that [the network shall] "have accounting mechanisms which allow users or groups of users to be charged for their usage of copyrighted materials available over the Network and, where appropriate and technically feasible, for their usage of the Network."[2] The predecessors for the NREN are NSFnet, a network connecting National Science Foundation-sponsored supercomputers to regional networks, and the Internet , an amalgam of linkages connecting universities, research organizations, military researchers, and government agencies. Both of these are successors to the ARPANET. One of the policies that is being implemented at the National Science Foundation is to allow access to NSFnet from commercial networks. Advanced Network Services, which manages the NSFnet, has set up a for-profit subsidiary called ANS CO+RE to sell access to computer networking. Other companies including Performance Systems International have set up a commercial alternative to the NSFnet called CIX and have protested that they do not have equal access to NSFnet since they do not get subsidized by the government.[3]

In the commercial marketplace services like TELENET and TYMNET developed that charge for use. From the earliest days of the ARPANET, the user had the perception that network use was a free service. NSFnet has an acceptable use policy which states that "use for commercial activities by for-profit institutions is generally not acceptable."[4] This paper will examine the implications for the academic community of the shift away from apparently free use toward a more fee-based network.

First of all, there are costs associated with networking. Organizations must interconnect their host computer to a network which is connected to NSFnet or the Internet, and must run software to package, route, and transport traffic. Users must have mailboxes and directories in which to store messages and files. It is the user and not the institution that thinks that the service is free. There is little doubt that the substantial costs of transporting and routing traffic through the network must be born by those who use the services. From an economic development perspective, if we really think that electronic exchange of information has value, it is in our best interests to make certain that funds to support it flow in so that expansion and improved service can be funded.

When we think of other communication services like the telephone and the mail, we have no assumption that usage will be free. We continue to use the telephone and the mail even though we realize that there will be a charge for their use. We are very likely to continue to use electronic mail, to access remote computers, and transfer files even though we realize there will be cost. With both telephone and mail there are subsidies that favor certain types of traffic -- rural traffic for telephone and nonprofit

or library traffic for the mail. The NREN is being set up to initially subsidize educational and research-oriented uses of the network, but this subsidy is unlikely to be complete or last indefinitely.

While organizations will have to pay to access and use the NREN, it is not clear to what extent they should extend charging to end users. First of all, charging mechanisms may be expensive to implement and administer, and even slight charges may have dramatic impacts on usage. People who would hesitate to call a 900 telephone number because of the cost feel no hesitation sending messages to hotlines all over the globe. Users may object that charging imposes censorship and brings commercialization into their traffic. If one thinks of telephone and mail use, few users pay charges out of their own pockets. Their organizations monitor charges and often step in if usage gets out of hand.

If we consider the economics of "free" services, we discover that non-price mechanisms develop for regulating use. For example, Internet users are quite familiar with overloads of traffic and difficulty getting help. Since there is no economic incentive to encourage use, techniques develop for discouraging use. These include lack of advertising, lack of training, lack of help, insensitivity to user needs, inadequate investment in upgrading facilities, lack of conversion of new ideas into marketable products, indifference to innovative thinking that might entail additional costs, and a general willingness to be contented with the status quo. I am not suggesting that the Internet or NREN does or will suffer from all these ailments, but it is probable that some will manifest themselves.

One of the hopes of many of us is that the United States will become a world leader in networked information services. In the late seventies I was involved in a study of the network information services marketplace headed by Herbert Dordick and Burt Nanus.[5] We felt that the United States had the potential through market forces to let industry decide which new services made economic sense and which did not. We would evolve from a number of private corporate networks to a public marketplace where information service vendors could carry on business with network users. NREN has many of the ingredients of the prototype of the public marketplace except that it does not have market forces and there is no mechanism in place to allow it to evolve into a market-oriented system.

A permeable boundary allowing traffic between commercial and nonprofit networks looks like a viable approach to allowing the commercial marketplace to grow up around the education and research community, but we cannot continue to prohibit commercial transactions on the NREN. It is inevitable that private sector organizations like publishers, equipment suppliers, and computer vendors will want access to their clientele. We will want to order books and software over the net and are likely to pay using our credit cards. What we need to make sure of is that the commercial traffic pays its own way and does not drive out the academic and research traffic. We also need to make certain that if high quality services develop in the commercial world, we can

access those services easily over the network even though there are free services available inside academia.

Some of the groups that have supported the NREN are represented by the Coalition for Networked Information. They have made sure that nonprofit and commercial databases will be made available over the NREN. Some people see the NREN as an escape from commercial publishers and commercial information utilities. They imagine free access and sharing of information as if there were no cost. It is a mistake to view the NREN as competing with the private sector; rather we should view it as carrying traffic that would not otherwise be picked up by the private sector, that has high social value, that serves the need of the disadvantaged, or that may involve high risk. There needs to be a smooth transition between what is on the research/education side of the network and what is on the commercial side so that services and products that can make it in the private sector can leave the public side and go commercial. Correspondingly, the education/research community needs to be encouraged to use private-sector services rather than insisting that these services be reinvented by the "free" network community.

What will it take to create smooth transitions? First of all, NREN users need cost information that will let them get some idea of how much various usages of the network are costing. They can use this feedback to understand the consequences of their actions, and can then make rational tradeoffs between the various media. They should also be given choices, so they can express their own priorities and values. For example, they ought to have the ability to request priority mail if it is essential, while requesting background file transfer for big files that are not needed immediately. There should be some mechanism for switching services if one is dissatisfied with the treatment one is receiving from the current network. If data transmission times are bad, or quality is deteriorating, one should be able to jump ship and go with another service. If a new vendor is offering an innovative service, it should be possible for them to advertise and steal away users from existing services. It should be possible for networks that really care about their users and who invest in training and user support to attract those users who think service is important. Correspondingly, those users who prefer to work things out for themselves should not be forced to pay for a network that emphasizes service.

As long as network usage is considered to be free, users and service providers are deprived of the opportunity to express their preferences and/or receive rewards for responding to unsatisfied needs. Enough people have now had experiences using networks to know that using them is valuable. It is time to let the developers of valuable services go private without losing access to their former users. Quality will have a chance of winning out in the marketplace only when users can vote with the dollars that come along with their usage.

1. 15 USC 5501 High Performance Computing Act of 1991

2. 15 USC 5512 sec. 102 (c) (6).

3. Markoff, John "U.S. Said to Play Favorites in Promoting Nationwide Computer Network N.Y. Times News Service 1991.

4. National Science Foundation. Interim NSFNET Acceptable Use Policy 6/14/90.

5. Dordick, H.S., Bradley, H.G., Nanus, B., Martin, T.H. "Network Information Services: The Emergence of an Industry" Telecommunications Policy 3(3):217-234 (Sept. 1979).

A COMPARISON OF COMMERCIAL E-MAIL PACKAGES AND SPECIAL INTEREST FORUMS FOR POSSIBLE USE BY MEMBERS OF PUPPETEERS OF AMERICA AND UNIMA AROUND THE WORLD

Gale Solotar Warshawsky
Lawrence Livermore National Laboratory

ABSTRACT: Communications technologies such as electronic bulletin boards, e-mail, and international networks are beginning to change the way people work and play, the way organizations conduct business and social interactions, and the very nature of how people communicate with each other regardless of where they live and work. Therefore, it becomes most important to be able to identify and select those systems which optimize the ability to carry out these information and communication needs.

1. INTRODUCTION: This paper is concerned with which commercial communication products are available to include E - Mail and Special Interest Forums for use by members of Puppeteers of America and UNIMA (*Union International de la Marionette*). There is a need for puppeteers in these organizations to be able to communicate with each other electronically no matter where in the world they happen to be performing. One of the factors that must be considered is cost. The membership of the Puppeteers of America is comprised of both professional and amateur puppeteers in the United States and Canada. The membership of UNIMA is made up of over 60 countries. Another factor to be considered is whether these commercial services are available in the countries with UNIMA members. Four commercial companies were researched to determine which would best meet the needs of the puppetry organizations. A comparative analysis, based on interviews with CompuServe, the WELL, GEnnie, and Prodigy was performed.

2. COMPUSERVE

CompuServe works on many different computers. I was most impressed with their literature, and by their helpfulness when I interviewed them over the telephone. CompuServe Mail lets users send brief messages and long documents to other CompuServe users and postal addresses, as well as MCI Mail, Telex, Internet, and fax users. This would be useful for puppeteers wanting to share scrips or articles with each other that they were co-authoring. CompuServe delivers the message and lets users know when they come online that they have mail waiting.

CompuServe has something called CB Simulator for conversation. CB has 72 channels or rooms to visit. Some groups are set aside for teenagers, others for adults or special support groups. Channels are available for spontaneous conversation, pre-arranged meetings, and private discussions. There is a National Bulletin Board for posting notices to the online community. Users may post items for sale, or browse want ads, job listings and special notices. Special features of CB Simulator permit users to carry on multiple conversations at the same time with people across town and around the world. I think this feature would be beneficial to the puppeteers. It would be handy for planning joint workshops to be held in a foreign country by various puppeteers from several different countries.

There are special interest forums for professionals and hobbyists online. Message boards are the most active place in a forum. Members can check it to catch up on the latest news, and to contribute to current discussions. Forum libraries include things like software files, professional newsletters, music scores, ham radio procedures, business plans, tax information, wine lists, gardening tips, etc. In all forums most library materials are free. Users may retrieve the file to their disk or add it to their personal resources. They can also contribute their own files for other members. This sounds like an excellent idea for the puppeteers. They want to be able to get information and add information electronically for everyone in the organizations to have access to. As CompuServe covers many countries world wide, this could give them the mechanism to communicate with each other conveniently. There are presently over 200 forums available.

This work represents the opinions of the author alone, and does not represent any work performed either by or for the Lawrence Livermore National Laboratory, the Department of Energy, or the United States Government.

The puppeteers could request that a puppetry forum be established for their use. A request would be made to CompuServe who would determine if that service would be worthwhile to them. All of the countries of the world except for 20 can access CompuServe. Connecting to CompuServe from a foreign country depends on the quality of the phone services available. In many Eastern Block countries there are poor local phone systems and many problems. CompuServe has access lists of local phone numbers in its member countries. European support is available from offices in Great Britain, Switzerland, and Germany. It would be hard to use CompuServe in some countries due to their phone line problems. For example, It takes 10 hours to place a phone call from East Germany, and that would affect using CompuServe. As a matter of fact, CompuServe is not available in East Germany at this time.[1]

The cost for the service is a one time membership fee of $39.95 which includes a User's Guide and a subscription to CompuServe's monthly magazine. Once you are a member you pay a basic connect-rate for the time you spend online. This rate is the same no matter what time of day or night you use CompuServe. The user supplies a personal computer, a modem, a telephone, and communications software. As an added bonus for signing up CompuServe gives new users $25.00 usage credit. This is about 2 hours of free time to explore CompuServe.

The connect time is billed in one-minute increments not including communications or premium product surcharges:

300 baud modem	$ 6.00/hr
1200 baud modem	$12.50/hr
9600 baud modem	$22.50/hr

Membership Support Fee $ 1.50/mo

Communication Surcharges:

To go online a uses dials CompuServe through their telephone and modem

Network	Prime	Standard
CompuServe	$ .30/hr	$ .30/hr
Data Pac (Canada Only)	$10.50/hr	$10.50/hr
800 Direct Access	$ 9.00/hr	$ 9.00/hr

Prime Hours are from 8:00 am to 7:00 pm weekdays.
Standard Hours are from 7:00 pm to 8:00 am weekdays and all day Saturday, Sunday, and specified holidays.[2]

Users can pay for monthly CompuServe charges with VISA, MasterCard or American Express credit cards. They can pay through CHECKFREE automatic electronic transfer from their checking account if offered by their bank. This is only available to members with US checking accounts and carries a $5.00 monthly minimum usage charge. Businesses with addresses in the US and Canada may apply for a business account. I inquired how a user without a charge card or US checking account could pay for the service. Many puppeteers do not have charge cards, but are legitimate business professionals. CompuServe told me they they could set up a personal business account and they would receive a monthly invoice which they could pay for by check.[3]

A user must supply their own computer, communications software, a modem and a telephone. CompuServe works on many different computers which include IBM, Macintosh, Apple II Series, Tandy, Atari, Commodore, and

Amiga. CompuServe supports 300, 1200, and 9600 baud rates on Hayes or Hayes compatible modems. When you join up for the service, they send you the software for your computer to access the service. [4]

A US puppeteer performing in a foreign country, i.e. Japan, could indeed hook up his laptop computer and modem and access CompuServe from that location. He would have access to E-Mail and the forum as well as CompuServe's other features. The only difficulty would be in making the local phone call from the foreign country's telecommunication's phone system. It is possible to print out E-Mail messages and forum information as well as download it to be saved to a disk.[5]

I think CompuServe would offer the members of Puppeteers of America and UNIMA the kind of online communication services they are interested in. I would recommend this company's product.

3. THE WELL

The WELL is a company whose users like to discuss technical things. The WELL (Whole Earth Electronic Link) also has the capability of connecting users up all over the world. They could create a puppetry special interest forum. It could be a private group, and there would be no charge to set it up. The WELL can be accessed from a variety of countries, and I was informed that the access depended upon if a particular country had a good telephone system. There is no support service in foreign countries. If users need help, they must dial the WELL which is based in California.[6] That would be an expensive phone call to get assistance as the telephone rates for calling the US from Europe are quite expensive.

The connections in Eastern Europe are very sparse. There is some action happening in the Soviet Union. Perhaps they could call into a node in Western Europe to get to the United States. Users in the Soviet Union, Yugoslavia, and Hungary can access the WELL.[7]

To use the service a user must have a computer, a modem, a telephone, and communications software. There is a $10.00/mo service charge and a $2.00/hr on line usage fee. The connect charges varies. It depends on point to point of where the call originates and terminates. It is possible to use packet networks to get cut rates for long distance phone calls. The preferred way to pay is by credit card. However, users may opt to pay a $25.00 processing fee to set up a billed account. An invoice would then be sent once a month and the user could pay by check. All new users get 5 hours free time. This service works on many different computers. It works with 300 baud, 1200 baud, and 2400 baud modems. It will soon work with 9600 baud modems as well.[8]

It would be very easy for a user to access the WELL and use its E-Mail and forum features from a country such as Japan, due to the advanced technology in that country. Users may print out E-Mail messages and forum items. They may also download these items onto their disk on their personal computer.[9]

I was impressed with the services of the WELL, however, not having local people in the foreign countries to assist users, necessitates long distance phone calls from the foreign country to California, US. This could be quite costly, and cause problems with the time differences. I don't think this company is the best choice for the members of Puppeteers of America or UNIMA.

4. GENNIE

GEnnie is owned by General Electric. It is a big packet switch time sharing network with 150,000 subscribers. Users could ask for a forum to be set up by talking with the manager of product marketing. GEnnie would add a puppetry forum to their hobby area. It could be a public or a private forum.[10] GEnnie is available in six countries as well as other cities which may be accessed by the local Packet Data Network (PDN).

In most countries of the world the government owns the PDN. Users get accounts with the PDN. Users receive 2 bills, one bill for local calls and one bill for GEnnie usage. Users pay by credit cards in countries other then the US and Canada. In the US and Canada users may use a checking account and pay by electronic funds transfer. If a user in one of the other countries that can access GEnnie does not have a credit card, they may not use GEnnie.[11]

The cost to use GEnnie is relatively small. It costs $4.95/mo which gives users access to 100 services which include E-Mail and the special interest forums. This rate is for the not prime time hours after 6 pm and before 8 am on business days, and all day on weekend and holidays. The cost in Canada is the same but users are billed in Canadian money. The cost in Europe and Japan depends on the local distributor or on the set price that GEnnie establishes within each country.[12]

GEnnie works on many different computers that can use communications software. Modems running at 300 baud, 1200 baud and 2400 baud have no fee attached to them for the 100 services. However, the 9600 baud modem will be added in limited locations and will have a $20.00/hr charge on top of the $4.95 GEnnie charge. Users need a computer, a modem, a telephone, and communication software to use GEnnie.[13]

If a US user wants to access GEnnie from another country outside of North America, they need to contact GEnnie ahead of time and clear it with them. They would have to tell GEnnie which country they would be in so that GEnnie could allow them to log into their system from that country.[14] Users may print out E-Mail messages and forum information as well as download this information onto their disk on their personal computer.

I don't think this company is the best choice for the members of Puppeteers of America or UNIMA. Although the service is very inexpensive, it does seem to have a drawback as far as user payment. If a user from a foreign country does not have a credit card then they may not use GEnnie. This company did not have any mechanism for setting up a business account for puppet companies who did not have a credit card and were located in a foreign country.

5. PRODIGY

Prodigy's service is only available in the US at this time. However, as the company is growing so rapidly, perhaps at some time in the future they might decide to go international. At this time, Prodigy is in 90% to 92% of the main cities in the US. Users may get up to 30 free messages a month by communicating with each other via E-Mail. Closed special interest groups may be set up. Prodigy has 800,000 members.[15]

Users pay for their service with one flat fee of $12.50/mo. If they desire to pay in advance they may pay $9.95 for 1 year's service or $8.33 for two year's service. Prodigy is available 21 1/2 hours a day for 30 days a month. If a user has more then the 30 free messages in E-Mail there is a $.25 charge per additional message. A user needs a computer and a modem to use Prodigy. They purchase the software in a Prodigy kit. This kit is available in several stores. The list price for it is $49.95 but it can be purchased at Sears for $39.95. New users receive one month's service for free. A user may cancel the service at any time by writing CANCEL across their Prodigy bill. Users pay by check. Credit card payment is not available.[16]

Prodigy only works on IBM personal computers, IBM clones, and Macintosh computers. It supports 1200 baud and 2400 baud Hayes or Hayes compatible modems. Prodigy suggests the 2400 baud modem because of the graphics used in their service. When a user purchases a Prodigy kit, they have an option of a kit with or without a modem.[17] (If a user needs a modem Prodigy will sell them one for approximately $100.00).

Users may print out E-Mail messages and items from the special interest groups. There are unlimited messages from the special interest groups at no extra charge. There is a large drawback of this service for puppeteers. There may not be any commercial messages on the special interest group's forums. This would indeed be a disadvantage to puppeteers. In the Puppetry Journal, one can find ads for various puppetry items of interest such as puppet supplies, puppetry books, and puppetry performances. Prodigy would not permit these types of messages to be included in the special interest forums as it would go against their no commercial messages rule. Prodigy does not support downloading from E-Mail or the special interest forums at this time.[18]

Prodigy is easy to use and can be hooked up and used very quickly. Up to 6 family members can use it, each having a separate password. It gives users a quick and easy access to information for the cost of a local phone call. There are no long distance phone call charges. It uses a graphics based interface and the graphics on it are pretty. However, it is not available outside of the US.

I don't think Prodigy would meet the needs of the members of Puppeteers of America and UNIMA at this time. However, I would recommend this service for puppeteers in the US who wanted to communicate with E-Mail with other puppeteers in the US. I don't think this would serve the membership's needs as far as special interest groups

forums are concerned. It seems like it is very easy to use, very inexpensive, and the customer service people were very nice to talk with.

6. CONCLUSIONS

A large concern I have for using E-Mail and special interest forums from the companies I researched include poor telephone systems in some countries. The success of connecting to these services is dependent on the quality of the telephone systems in the various counties. If a user is in Japan then there is no problem as Japan has excellent technology. However, if a user is in one of the Eastern Block countries, where it takes many hours to make a phone call, it won't matter how great the communication's company is. If the phone service and telecommunications of an area are poor, then the service offered by companies such as CompuServe, The WELL, or GEnnie won't do a user much good. An other concern is that many members of Puppeteers of America and UNIMA do not own computers. I think the idea of connecting the membership up electronically is a good one. Perhaps the members of these organizations could write grants that would convince Apple computer or IBM to give them computers. Some of the members are teachers. Apple has a program to get their computers into the public schools. This is an area that will need to be investigated by the membership if they are interested in pursuing E-Mail and special interest forums online. I think the need is there. Now all I have to do is convince the membership! Perhaps by the time the countries with poor telecommunications services modernize their phone systems, members of these puppetry organizations will have computers to use and can become part of the information age.

7. BIBLIOGRAPHY

Customer Service. CompuServe,Telephone Interview. 24 February 1991.

Harris Neil. GEnnie, Telephone Interview. 27 February 1991.

Hops, Al. Prodigy, Telephone Interview. 25 February 1991.

Questionnaire 23 February 1991.

Rhine, Nancy. The WELL, Telephone Interview. 25 February 1991.

Sterling, Christopher. *International Telecommunications and Information Policy.* Communications Press, Inc.

Washington DC, 1984. 496 p.

8. REFERENCES

1 CompuServe Customer Service , telephone interview on February 25, 1991.
2 CompuServe Information Sheets
3 CompuServe Customer Service, telephone interview on February 25, 1991.
4 *Ibid*, February 25, 1991.
5 *Ibid*, February 25, 1991.
6 Nancy Rhine, The WELL, telephone interview, February 25, 1991.
7 *Ibid*, February 25, 1991.
8 *Ibid*, February 25, 1991.
9 *Ibid*, February 25, 1991.
10 Neil Harris, telephone interview, February 27, 1991.
11 *Ibid*, February 27, 1991.
12 *Ibid*, February 27, 1991.
13 *Ibid*, February 27, 1991.
14 *Ibid*, February 27, 1991.
15 Al Hops, Prodigy, telephone interview, February 25, 1991.
16 *Ibid*, February 25, 1991.
17 *Ibid*, February 25, 1991.
18 *Ibid*, February 25, 1991.

EXTENDED ABSTRACT

Rights, Roles, Rules: Some Ethical Concerns for Academics Using Electronic Discussion Groups

Robin Peek
Graduate School of Library and Information Science, Simmons College
Boston, MA

Electronic Discussion Groups (EDGs) which serve as "public" forms using LISTSERV mailing "lists" on BITNET (or USENET newsgroups, among others) are becoming more popular with scholars in the social sciences and the humanities as a means of discourse about professional interests, as is evidenced by the recent growth in the number of these groups available. It is important that now, in this early period of the development of these EDGs and their use by scholars in these disciplines, that the patterns of use and fundamental tensions that are emerging be identified as it will influence the future structure and patterns of these groups. Central to this concern is decision-making used by the EDG "owner" or "editor" regarding the control or lack of control of these messages as they are distributed to the group membership.

This study initially examined nineteen scholar-focused, open or "public", EDGs with broad topical interests in the social sciences and humanities that use the LISTSERV mailing list feature on BITNET. From this pool nine groups were selected that remained "active" with regular message transmission over a six month period. Content analysis was conducted on this pool of data. Semi-structured interviews were then conducted with 30 members of these groups. Each respondent was either a full-time or part-time faculty member at a college or university. From this data, a pattern of concerns regarding the management of these groups emerged. It is the purpose of this paper to summarize these findings around the central theme of the issues of the role of group "owner", "moderator" or "editor" (hereafter referred to as EDG owner) to other group members. This paper describes the norms of conduct that concern EDG participants (the members of the EDG) about the "if", "how", and "who" that control the flow of information that is transmitted. It is not the purpose of this paper to establish a rigid ethical framework, however, it is to suggest that there are areas of ethical concern regarding decision-making made by the owner that impact on the nature of the communications received by the membership.

In discussing roles, rules, and responsibilities of the EDG owner in a scholarly-oriented group, the function is an evolving one. Social scientists and humanists are still relative newcomers in this medium and bring their own norms and values from both their professional orientation and from the tradition of scholarly communication. The EDGs that exist on BITNET are in a unique environment, different from that of other networks, because BITNET primarily serves colleges and universities, unlike other loosely-organized forums such as USENET that serve a broader and more diverse audience. Thus, as the scholarly audience from these disciplines interact with each other in this environment, the role of the EDG owner, who has the first and perhaps final say regarding EDG conduct, has a particularly important role.

In order for a group to exist on BITNET it must be established and "owned" by an individual or a group of individuals. It should be noted though that ownership can and often does change throughout the lifecycle of the group. The policy that directs which individuals can be EDG owners varies from institution to institution depending upon the policies established by the campus computing center. Because of this an owner can be a graduate student at one institution, a staff person at another, or

a professor at yet another. Many of the groups are co-owned by two or more individuals.

Regardless of how a group is created, the decision to own a group is a voluntary one that normally provides little or no reward in terms of promotion, tenure, or money. This must underlie one of first decisions that must be made, and that is the involvement of the owner in the exchange of messages within the group as they pass through the listserver. Utilizing terminology that evolved during the early days of these groups in general, a group is typically known as being either "moderated", where the list owner injects varying degrees of control over what is actually posted to the group and how it is arranged, or it is "unmoderated" where essentially everything that is sent to this group is posted "as is". In the unmoderated group the list owner does not see the messages before they are posted to the EDG.

Owning an EDG can be a time-consuming proposition, particularly on a moderated group that has heavy message traffic. Sifting through incoming messages can easily become a daily chore that can take upwards of an hour or more a day. It is possible, as in the case of one EDG, that this chore be handled by a graduate assistant, but this is by far the exception and not the rule. This issue of time, of course, creates a fundamental problem in any call to improve the quality of discourse by requesting a more active role on the part of the list owner.

How much time the management of these groups takes depends both on the message traffic from the group and the degree of involvement the owner decides to inject into the functioning of the group. Whereas some moderators will send messages that they have not decided to put on the EDG back to the original sender with a message noting why the message was rejected and perhaps making a recommendation for rewording or suggesting another EDG that may be more appropriate, other moderators will merely not post the message. Moderators may also elect to involve themselves actively in the group message stream by cross-posting messages from other groups, posting questions to stimulate exchange or controlling the direction of discussion through their own posting of messages. The amount of involvement that an EDG can elect to take may be quite high. Moderators inject editorial decisions by winnowing out messages that they deem inappropriate to the group.

But The role of a EDG owner is ill-defined particularly within the realm of exchanges between scholars. While scholarly communication has evolved a set of traditions regarding the role of the editor of a traditional print publication or a chair of a conference meeting which is generally understood by all of the players involved, there is currently no norm for an EDG. Therefore, the moderator of an electronic discussion group is far less understood by both the moderators themselves and by the people who participate in these groups.

Call for More Structure of EDGs

While acknowledging that there is a place for unmoderated groups (although they may not choose to participate in them), most scholars found that the more moderated, structured, and narrowly focused in topic EDGs were a more preferable form of discourse and predicted this as the future trend of these EDGs. This is not surprising as it parellels traditional scholarly communication. But this then places more emphasis on the decision-making of the EDG owner in how they create the group and to what extent they control it.

This returns us to the central question, what are the roles, rules, and responsiblities that guide the behavior of present and future EDG owners in an environment where there is little tangiable reward for their efforts. There was indeed debate among the respondents themselves as to what extent, if any, the EDG owners should be rewarded through the promotion and tenure system for the work required to moderate a group. A futher problem is

posed by the fact that while a member may elect to participate or not in a group, it is quite difficult to excise the EDG owner, since it is a self-selected role.

Given these constraints, what can an EDG owner do to enhance the relationship between the owner and the membership? A few suggestions emerged from this study and reflect more common courtesy than rigid guidelines.

1) An EDG owner is obligated to not abandon the group--to leave it unowned. The owner should also read the messages whether the group is unmoderated or moderated. If the owner chooses not to continue in the role, the owner is obligated to close the group down if a replacement owner cannot be found.

2) The EDG owner should send new members an introductory message that outlines the purposes of the group and whether it is unmoderated or moderated. If a certain type of discourse is not welcomed, it should be stated openly (i.e. long discussions on the EDG, cross-postings, etc). The owner should be clearly identified. If ownership changes, that needs to be communicated to the group.

3) This message should be sent out to the membership at regular intervals to remind the membership of the intent of the EDG. If the intent changes, the group should be informed of that as well.

4) The owner should also regularly post a message to the group that explains technical aspects of the EDG such as leaving the group and posting a message. On an unmoderated group in particular this would reduce the number of error postings.

5) If the status of a group changes from an unmoderated environment to a moderated environment or vice versa, the EDG owner should inform the group of this decision.

6) If the EDG owner elects to require the filling in of an application prior to that person being allowed to participate in that group, the rationale for this decision and the selection criteria for admission should be publicly stated when the application is sent to the potential member.

7) If a posting is rejected for the EDG, the list owner should return the message to the sender with an explanation. While this could be time-consuming, it would reduce the tensions created between the owner and the participant when a message is not posted.

Networked Resource Sharing of CD-ROM Information Banks

Fredric C. Gey
UC Data Archive & Technical Assistance (UC DATA)
University of California, Berkeley
e-mail: gey@sdp2.berkeley.edu

ABSTRACT

1991 was the year in which the CD-ROM finally arrived as a publishing medium. The federal government published over 500 CD-ROMs in calender year 1991, increasing the total number of publically available CD-ROMs by more than one third. Unfortunately, the primary mode of usage of CD-ROM databases has each single user occupying a personal computer (with CD-ROM hardware attached) for the duration of her research, leaving other researchers waiting until the resource can become available. Clearly, this paradigm is inadequate for practical support of this proliferation of CD-ROMs by a multiplicity of researchers and an inquisitive student body.

This paper proposes an academic information technology infrastructure to deliver information 'where the action is,' i.e. to the individual student, librarian, or academic researcher. The vehicle is a distributed network environment making available CD-ROM jukebox devices which hold up to 30 CD-ROMs per service workstation. Operation during limited hours is replaced with open 24 hour access to data from any network-connected UNIX or IBM-PC-compatible workstation. Given an appropriate campus network one can ultimately envision undergraduates accessing data from their dormitory rooms over the campus network to the information server. This prototype facility could be duplicated at modest expense at any campus (or even within academic departments and individual library units).

1. Introduction

The UC Data Archive and Technical Assistance (UC DATA) supplies and supports quantitative social science and health statistics databases for the UC Berkeley campus. Until recently most of these databases have been delivered from supplying organizations on mainframe computer tape. The user community for this form of data has, for the most part, consisted of faculty researchers and advanced graduate students skilled in the use of mainframe statistical analysis software such as SPSS or SAS.

During the past 12 months UC DATA has received more than one hundred CD-ROMs of social science data, mainly from the Census Bureau. With the impending release of numerous additional CD-ROMs from the 1990 census, this number should grow to significantly by the end of 1992. One hundred CD-ROMS is equivalent to four hundred twelve-inch magnetic tapes. Thus in a single year, the distribution of data in CD-ROM form has increased our holdings by more than 10 percent, without additional

staffing or resources to deal with this information explosion. This situation is faced by government document libraries and social science information centers throughout the country.

The primary bottleneck the academic social science community faces in using this information is the single-user, single task personal computer typically used to access the data. If a researcher or student actually has such a PC with the CD-ROM hardware attached, then considerable work can be accomplished on a single database at a time. Moreover, the Census Bureau has provided user-friendly profile software which will produce reports on specific geographic areas. Figure 1 shows such a profile from the 1988 County and City Data Book for Austin, Texas.

```
                    COUNTY AND CITY DATA BOOK 1988

 STATES                                 Cities of 25,000 or more
 AL AK AZ AR CA CO CT DE DC FL          State Totals
 GA HI ID IL IN IA KS KY LA ME          Abilene
 MD MA MI MN MS MO MT NE NV NH          Amarillo
 NJ NM NY NC ND OH OK OR PA RI          Arlington
 SC SD TN TX UT VT VA WA WV WI          Austin
 WY  TEXAS                              Baytown

 SUBJECTS  Land Area and Population
 LAND AREA, 1985  (SQUARE MILES)..............................     232.0
 LAND AREA, 1980  (SQUARE MILES)..............................     116.0
 TOTAL PERSONS, 1986..........................................   466,550
 RANK OF CITY POPULATION, 1986................................        27
 PERSONS PER SQUARE MILE, 1986................................     2,011
 TOTAL PERSONS, 1980 (CORRECTED)..............................   345,890
 NET CHANGE, 1980-1986........................................   120,660
 PERCENT CHANGE, 1980-1986....................................      34.9
 POPULATION CHARACTERISTICS, 1980:
   PERCENT WHITE..............................................      75.6
   PERCENT BLACK..............................................      12.2
   PERCENT AMERICAN INDIAN, ESKIMO, AND ALEUT.................       0.3

 ↑↓-Cursor  P-Print   PgUp-Page Up  PgDn-Page Down  Esc-Reset  F1-Flag Legend
```

Figure 1: County City Data Book for Austin Texas

The availability of such software means that social science undergraduates without significant computer expertise could, in principle, access this data and utilize it for course projects. Indeed, as UC DATA has entered its machine-readable data collection into the on-line book catalogs of the Berkeley campus and the University of California nine-campus catalog (MELVYL), our educational service clientele has expanded from faculty and advanced graduate students to include undergraduates who have learned about us from the catalogs.

However, if you multiply the desired access by the potentially thousands of undergraduate users on any major university campus, the existing resources (a few PC available during limited hours of operation) is not up to the task. Moreover, adding more PCs is a limited solution; a more radical approach is called for.

This paper develops a prototype solution to the access problem such that data can be

made available to inquiring students and faculty researchers without leaving their offices or their departmental computing laboratories.

2. Characteristics of government numeric data on CD-ROM

In contrast to the bibligraphic and textual databases most often found in the CD-ROM medium in libraries and information centers, the government statistical information has different access and format characteristics. The data is primarily geographic in nature, and access is focused on the individual geographic unit, be it state, county, census tract, census block group, or even city block. Examples of such databases include the following:

- *1990 Census Summary Tape File 1 (STF1)* -- These eight CD-ROMs contain the latest available information from the 1990 Census. Over 300,000 records for census block groups (a unit of area defined by the Census Bureau as comprising about 250 households or about 1000 individuals), for the entire United States. Of particular interest are age-race-sex distributions, and median rents and median values of owner-occupied housing.

- *1988 County and City Data Book* -- This database contains the most comprehensive cross-section of information for each county in the U.S., all cities of 25,000 or greater population, and several items (population and per-capita income) for all towns of 2,500 or greater population. The county and city files include information on vital statistics, crime, agriculture, government finances, economic activity, and decennial census information.

- *Bureau of Economic Analysis Regional Economic Information System* -- The Bureau of Economic Analysis CD-ROM consolidates a time series of Local Area Personal Income from 1969-1988. This important county-level database includes segments of income derived from non-wage sources such as pensions, interest and dividends, and cash and non-cash benefit programs.

From such examples we can deduce some general characteristics which contrast such numeric data from the usual library fare:

- *Data is public domain* -- so issues of copyright infringement and licensing for multiple or networked access do not exist. In particular, information on the CD-ROM can be copied to faster magnetic disk media when warranted for improved multi-user access speeds.

- *Databases often extend across multi-CD sets,* making the volume of information considerably larger. However, by the same token,

- *Access patterns are intermittent and localized,* since the user of the data is generally searching for demographic patterns within small communities or across counties. This means that CD-ROM towers with a drive per disk may not be necessary,

and juke boxes with one drive for many CD disks may serve adequately.

- *Data storage is standardized,* and hence information can be accessed by using standard database systems such as DBASE-IV or programming languages such as BASIC and C.

- *Bundled access software is vanilla,* and hence often doesn't require extra memory or utilize fancy screen addressing techniques which might break down when moved to UNIX workstations with PC-DOS emulation software.

These characteristics argue against the use of expensive proprietary access software which is tied to a limited selection of particular databases or to particular hardware services optimized for such software and database combinations.

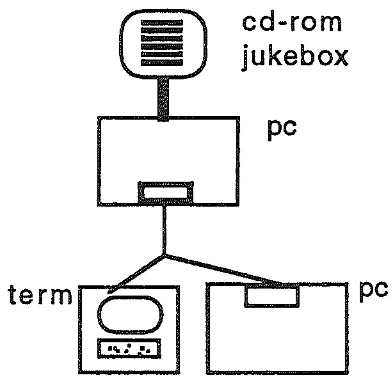
3. Networked solutions to the CD-ROM access bottleneck

Access and utilization of CD-ROMs over computer networks has been operational in few organizations within the past two years, although most libraries are considering such an installation to cope with the increasing popularity of CD-ROM information. A recent survey of academic libraries noted only 21 networks in 77 institutions responding [LaHuDo 91]. Almost all such installations described in the literature utilize PC-based Local Area Network products. [PeThGu 91] describes alternative configurations and evaluates products which implement such CD-ROM IBM-PC LAN-based environments.

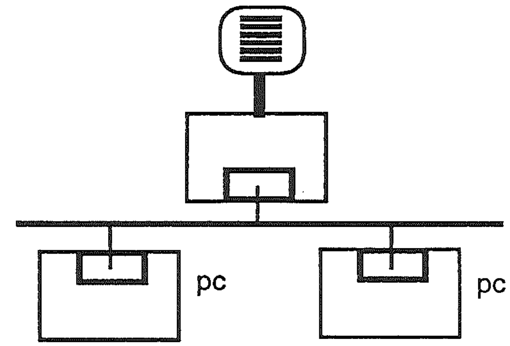
Figure 2 shows several possible simple-to-complex network solutions to the CD-ROM networking problem. First (2a) is a program such as PC-ANYWHERE which acts similarly to a bulletin board service by allowing terminal or terminal emulation access to programs on the PC which runs the CD-ROM software. This solution is cheap but quickly runs into the same barrier of single-tasking machine access on a PC. Second is to use *redirection software* (2b) which allows networked PCs to access another (peer) PC which has the CD-ROM attached, making it look to the accessing PC as if the CD-ROM were locally attached. The disadvantage is that the Microsoft CD-ROM extensions (driver software) must be installed on each accessing PC..

A third option (2c), and the one in common use on LANs around the country, attaches a *CD-ROM Server* to the LAN and accessing PCs transparently use the CD-ROM without special drivers. Often the Server configuration includes multi-drive CD-ROM Towers with up to 12 drives, one per CD and special software on the Server machine to cache requests and data and thus enhance the speed of access. These Towers, although providing a drive for each CD-ROM mounted, and dedicated CD-ROM server hardware on a PC with substantial main memory, and thus providing multiple simultaneous access at reasonable speed, are very expensive.

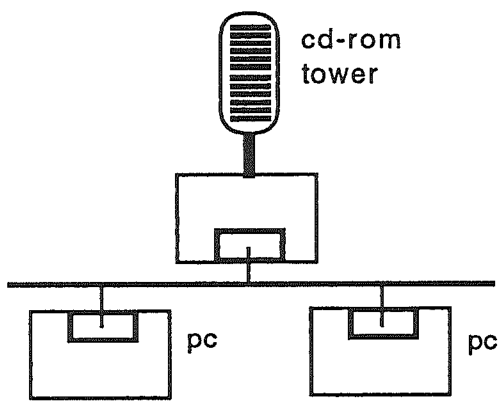
A somewhat different option (2d) is available using the V/Server from Virtual Microsystems which is a circuit board installable on a Digital Equipment Corporation



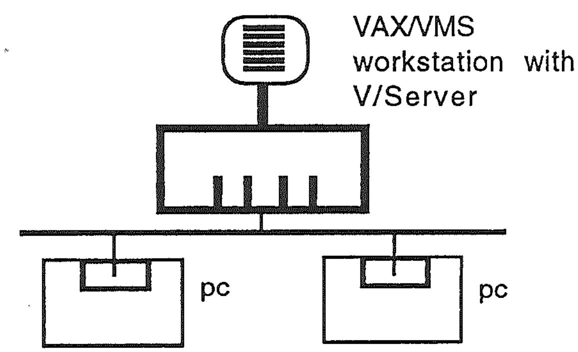
2a) PC with terminal session software



2b) PCs with redirection software



2c) PC -LAN with CD-ROM Server



2d) LAN with VAX and V/Server

Figure 2: CD-ROM Network Alternatives

VAX computer running the VMS operating system; this board has four PC processors capable of handling PC software and making CD-ROM data available as if it were coming from a PC. This solution, which also enables the running of PC software as tasks on the VAX computer, is effective, yet expensive. The above options are described in somewhat more detail in [KOREN 91] and [JaWa 92].

Notably absent from such evaluations are alternatives which attach CD-ROMs to a UNIX workstation which would make the resource available to the campus-wide networks (indeed, to the world-wide InterNet) running the TCP/IP protocol.

3.1. Desired characteristics of networked CD-ROM access

Rather than approach the implementation of multi-user CD-ROM access from the point of view of which products are available to implement such access, perhaps a better starting point is to identify the desired characteristics which would maximize networked CD-ROM access. Certainly the following features would be desirable:

- *Access should be available from any point* on a campus or organization-wide network, not merely from a LAN of limited geographic scope.

- *CD-ROM databases should be available from multiple servers* enabling components of a CD-ROM information bank to be geographically dispersed with each component under local control of experts most familiar with the data content.

- *Access should be available from diverse hardware configurations* and operating systems including, but not limited to, PCs, MacIntoshes, UNIX workstations, and even IBM Mainframe computers.

- *Configurations should support multiple access modes* including direct disk access to the CD-ROM data from any client PC or workstation, as well as terminal session access for users without PCs or from a window on a remote workstation.

- *Hierarchies of electronic storage should be supported* by the configuration so that popular databases experiencing heavy demand can be moved from slower CD-ROM media to standard magnetic disk media capable of supporting such demand.

- *The architecture should support incremental evolutionary expansion* inexpensively, whether in the form of additional CD-ROM capacity, or when adding additional servers to the configuration.

All the above criteria argue against the traditional CD-ROM services configurations as described in [PeThGu 91] and in favor of a new solution utilizing UNIX workstation-based capabilities. UNIX Network File Services (NFS) software enables multiple nodes on a network to act as disk servers for all other workstations, allowing the distribution of CD-ROM databases to multiple geographic locations. In addition, utilization of PC-NFS software will allow PCs to remotely mount disks over the network

(unrestricted as to location) if given permission by the serving workstation. This would integrate PCs into the environment in a similar yet more general way than with proprietary, dedicated PC-based LANs.

4. Hardware and Network Configuration

Figure 3 (on the following page) displays a diagram of the planned configuration of the project's hardware and network components. The SUN SPARC workstation will reside at UC DATA facilities on a Local Area Network (LAN). Program. Daisy-chained to the workstation through a SCSI port will be 5 Pioneer DRM-910 CD-ROM Changers (jukeboxes) each holding 6 CD disks in a removable caddy. Also attached to the LAN will be two PCs (or compatibles) with ethernet LAN connection, running SUN's PC-NFS software which enables them to directly address the CD-ROMs as if they were locally attached to the PCs. Equivalent PCs will be connected from the UC Library's Government Documents section, the School of Library and Information Studies (SLIS) Bibliography Laboratory, and the Political Science Department's Computing Laboratory. Direct access to CD-ROM data by UNIX workstations at the Quantitative Anthropology Laboratory (QAL), the School of Library and Information Studies, and the Survey Research Center will also be provided from the UC DATA workstation.

5. Prototype

An installation similar to the proposed configuration, using a SUN Microsystems SPARC-1 workstation with 3 Pioneer Jukeboxes attached, has been operational at Lawrence Berkeley Laboratory* for the past six months. UC DATA, using an ethernet card and PC-NFS software has, with the permission of the LBL investigator, accessed this installation from our PC located just south of the Berkeley campus. Reasonable access speed has been achieved on these remotely mounted CD-ROM drives which are physically located about a mile away. The current access time seems to be limited by the 59Kbaud network connection from the UC DATA ethernet to the campus network.

6. Database installation and usage

The availability and variety of databases on CD-ROM depends largely upon who is using CD-ROM as a publishing medium. Since the Bureau of the Census has been a leader in converting to CD-ROM, the databases for initial installation will be heavily weighted toward Census Bureau disks. In the course of the project we can expect a wider variety of social science databases to appear in the CD-ROM medium. The extended life-cycle of an information access project depends upon provision for training and educational materials to instruct in usage of the information bank. Components of instruction on usage by faculty researchers and graduate students will

*Deane W. Merrill, Biostatistics Program, Information and Computing Science Division, Lawrence Berkeley Laboratory, private communication

121

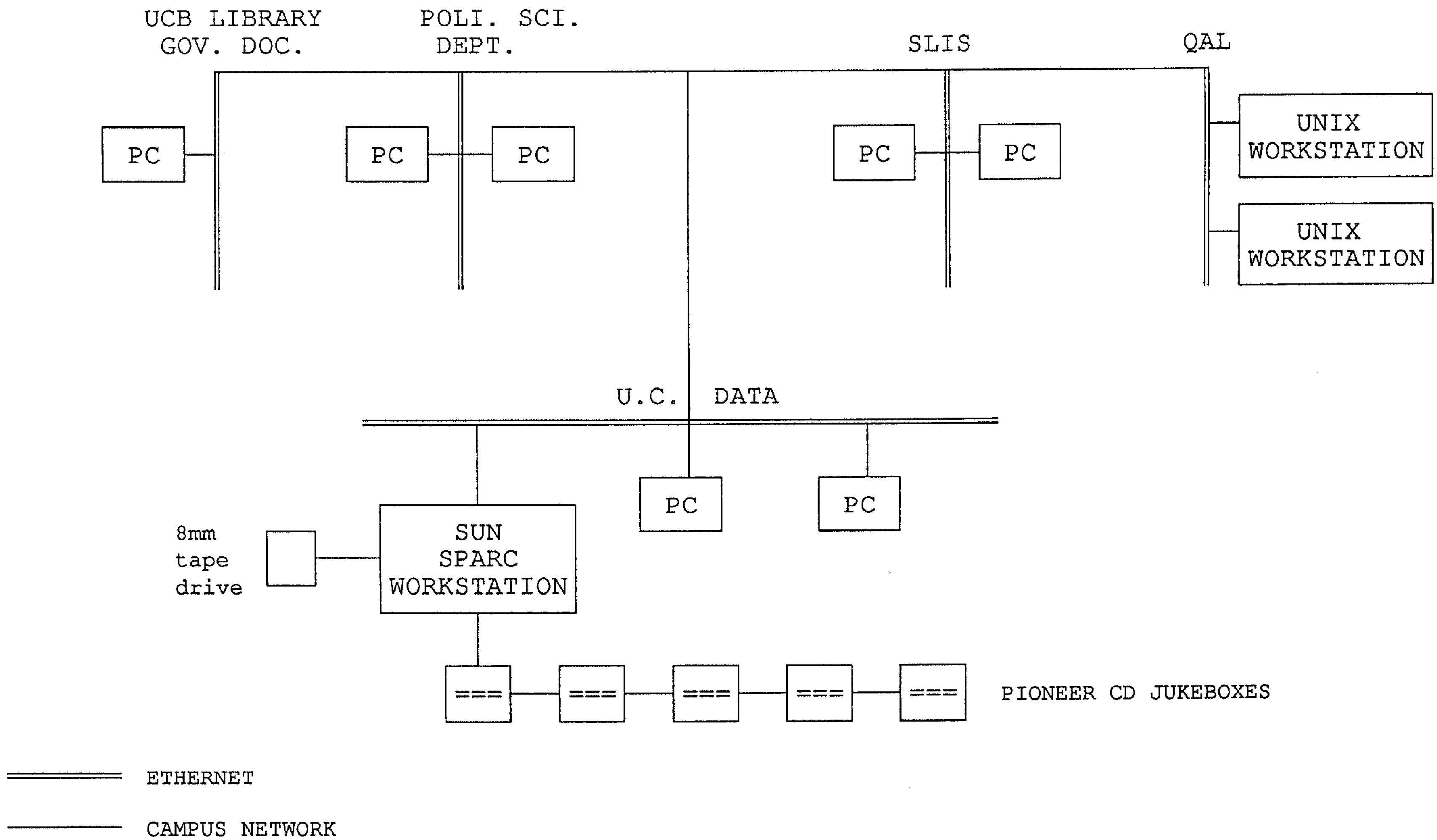


Figure 3: Project Hardware and Network Configuration

include:

- *Finding the data to be analyzed.* On-line subject searching of individual data elements is beyond the scope of this proposal. UC DATA will prepare a directory of data sets installed on CD-ROM and make available the detailed technical documentation for each database.

- *Software to access data.* Which pieces of software are available to profile, extract, tabulate, and otherwise manipulate the information?

The easy way for novice computer users to access the data is to utilize user-friendly profile software such as displayed in Figure 1 above. A somewhat more challenging approach is for users to learn the intricacies of general purpose (but menu-driven) software such as the Census Bureau's EXTRACT program. These programs allow on-line selection of fields in the database, as well as selection of groups of records. Capabilities for output formatting and selection of types of output file formats for data extracted in machine-readable form (e.g. DBASE, SAS, Lotus worksheet) are also included. The most difficult approach is for the researcher or student to learn to use general purpose software such as DBASE, SPSS, or SAS. DBASE is particularly important with Census databases, since the Census Bureau has standardized on the DBASE-III file format for the distribution of their data on CD-ROM.

7. Summary

This paper has described the desired configuration requirements and database characteristics to provide organization-wide access to government statistical databases issued on CD-ROM. The requirements suggest that the usual approach of information services based upon PC-LANs will not suffice for extensive CD-ROM collections, and heterogeneous access needs. A CD-ROM jukebox configuration attached to UNIX workstations offers the greatest flexibility of both storage and access with the least incremental expansion overhead.

8. References and Bibliography

DESM 91 — Desmarais, Norman, "Easy Access to Multiple CD-ROMs," *CD-ROM Librarian*, V. 6, No. 2, February 1991, pp23-24.

JaWa 92 — Jaffe, Lee David and Steven G. Watkins "CD-ROM Hardware Configurations: Selection and Design" *CD-ROM Professional*, January 1992, pp62-68.

KOREN 91 — Koren, Judy, "CD-ROMS and CD-ROM Networks, and their Connection to the Campus (Ethernet or DECnet) Network," Special Report for the Library, Technion University, Hiafa, Israel, June 12, 1991, distributed electronically via the CDROM-LAN ListServer.

KrJaAr 91 — Kriz, Harry M., Nikhil Jain and E. Alan Armstrong "An Environmental Approach to CD-ROM Networking using Off-The-Shelf Components," *CD-ROM Professional*, July 1991, pp24-31.

LaHuDo 91 LaGuardia, Cheryl, Chuck Huber, and Carol Doyle, "CD-ROM Networking in ARL Academic Libraries: A Survey," *CD-ROM Professional,* March, 1991, pp36-39.

MASS 90 Massey-Burzio, Virginia, "The Multiplatter Experience at Brandeis University," *CD-ROM Professional,* May 1990, pp22-26.

NEWM 91 Newman, Linda, "Building a CD-ROM Local Area Network: Life on the Bleeding Edge," Southern Ohio Chapter of ASIS workshop: *Networking Databases: CD-ROM vs. Mainframe,* Cincinatti, Ohio, March 22, 1991.

PeThGu 91 Perratore, Ed, M. Keith Thompson, and Angela Gunn, "Networking CD-ROMs," *PC Magazine,* December 31, 1991, pp. 333-363.

PERR 90 Perry, Edwin M, "The Ezekiel Effect -- Factors Affecting Development of Local Area Networks for CD-ROM," *Laserdisk Professional,* January 1990, pp7-9.

ThMa 90 Thompson, M. Keith, and Kimberly Maxwell, "Networking CD-ROMs," *PC Magazine,* 2/27/90, pp237-260.

Interfaces for Distributed Systems of Information Servers

3/1/92 Version 1.0

Brewster Kahle, Harry Morris, Jonathan Goldman
Thinking Machines Corporation

Thomas Erickson
Apple Computer

John Curran
NSF Network Service Center

Email addresses: brewster@think.com, morris@think.com, jonathan@think.com, thomas@apple.com, jcurran@nnsc.nsf.net

ABSTRACT

Interfaces for information access and retrieval are a long way from the ideal of the electronic book that you can cuddle up with in bed. Nevertheless, today's interfaces are coming closer to supporting browsing, selection, and retrieval of remote information by non-technical users.

This paper describes 5 interfaces to distributed systems of servers that have been designed and implemented: WAIStation for the Macintosh, XWAIS for X Windows, GWAIS for Gnu-Emacs, SWAIS for dumb terminals, and Rosebud for the Macintosh. These interfaces talk to one of two server systems: the Wide Area Information Server (WAIS) system on the internet, and the Rosebud Server System, on an internal network at Apple Computer. Both server systems are built on Z39.50, a standard protocol, and thus support access to a wide range of remote databases.

The interfaces described here reflect a variety of design constraints. Such constraints range from the mundane—coping with dumb terminals and limited screen space—to the challenging. Among the challenges addressed are how to provide passive alerts, how to make information easily scannable, and how to support retrieval and browsing by non-technical users. There are a variety of other issues which have received little or no attention, including budgeting money for access to 'for pay' databases, privacy, and how to assist users in finding out which of a large (changing) set of databases holds relevant information. We hope that the challenges we have identified, as well as the existence and public availability of source code for the WAIS system, will serve as a stimulus for further design work on interfaces for information retrieval.

1. INTRODUCTION

It requires little prescience to predict that one day computers will put an ocean of information at the finger tips of a vast population of users. However, although there is a considerable amount of information available from remote sources, the bulk of it is accessible only to information professionals, or users with technical backgrounds. A variety of obstacles effectively block the ordinary user from accessing information via the computer. These obstacles include the difficulty of locating appropriate information sources, the cumbersome maneuvers needed to get on-line and to connect to remote sources, and cryptic query languages. Furthermore, even if a user has succeeded in accessing a remote information source, it is likely that it will have its own special purpose interface, which may or may not support the user's needs.

In this paper we describe two systems—Wide Area Information Servers (WAIS), and Rosebud—which provide a protocol-based mechanism for accessing a variety of remote, full text information servers. These systems have the potential for supporting a single interface to a wide variety of information sources, and offer a good platform on which to explore the design of interfaces for information retrieval. After a summary of existing information retrieval systems, we describe the server systems, and then describe the 5 interfaces to them. In the course of these descriptions we discuss design constraints, interface issues, and practical matters which impacted the designs. We conclude with a summary, and some remarks on important issues which have not been addressed, and a invitation for other investigators to use the WAIS system as a platform for exploring interfaces to multiple, remote information sources.

2. BACKGROUND

2.1 Existing Systems

While a review of all existing systems is beyond the scope of this paper, it is useful to list a number of the most popular or significant interfaces for information retrieval.

Commercial interfaces for accessing full text resources on computers can be broken down into dialup services, local file access, and LAN-based access tools. Dialup systems such as Dialog and Dow Jones offer TTY interfaces to users, with menus and command lines being the dominant access tools. Some dialup services are offering client programs that run on personal computers to add graphical interfaces such as "Navigator" by Compuserve. In general, these interfaces are unique to the information provider. Local file access through full-text indexing has been achieved in command line form (e.g. the unix command "grep") and in screen based interfaces (e.g. ON Location (ON), and Digital Librarian (NeXT)). These interfaces often give browsing and searching capabilities for local files. Some of these interfaces have been stretched to work with files on file servers. LAN-based access tools usually use some sort of query language to access servers on the net, such as Verity's Topic system (VERITY), and numerous library systems. These query languages require some user training. Integrated tools for cross platform, cross vendor information access are not currently available in other systems.

A variety of research projects have explored information retrieval systems. The SuperBook project (Egan, 1989) targets users of static information. Project Mercury (Ginther-Webster, 1990) is a remote library searching system that uses a client-server model. Information Lens (Malone, 1986) is a structured email system for assisting in managing corporate information. NetLib for software (Dongarra, 1987) and Mosis for information on how to fabricate chips (Mosis) are examples of email based information retrieval systems.

2.2 The WAIS and Rosebud Projects

The two systems of information servers described in this paper grew out of two, partially entwined projects: WAIS, and Rosebud. A goal of both projects was to define an open protocol that would allow any user interface or information server that talked the protocol to interact with any other component which used the protocol. From the user's perspective, this would mean that user interfaces and information sources could be mixed and matched, according to the user's needs.

WAIS started as a joint project between Thinking Machines Corporation, Apple Computer, Dow Jones & Co., and KPMG Peat Marwick (Kahle, 1991a). The proximate goal was to define the open protocol and demonstrate its

feasibility by implementing and demonstrating a multi-vendor system which provided ordinary users with access to a variety of remote databases. Thinking Machines contributed its Connection Machine based retrieval technology, Apple contributed its expertise in user studies and interface design, and Dow Jones & Co. provided access to its commercial information sources. KPMG Peat Marwick provided access to its corporate data, and served as a site for user studies and testing. The WAIS system was installed at KPMG Peat Marwick and enabled the designers to study the success of the system in a real world context. The WAIS system uses pseudo natural language queries, relevance feedback to refine queries, and accesses full text, unstructured information sources. These technologies were used because they had already been tested independently, thereby leading to faster implementation of the complete system. The WAIS system will be described in more detail in the next section.

During the same period, the Rosebud project was underway within Apple. Rosebud's goal was to serve as an internal platform for research into system architecture and human interface issues, and as a consequence employed a variety of more experimental technologies and was tested in-house. Like WAIS, Rosebud was based on user studies conducted at KPMG Peat Marwick, and used the same underlying protocol, Z39.50. The details of the Rosebud Server System will be described in a different paper.

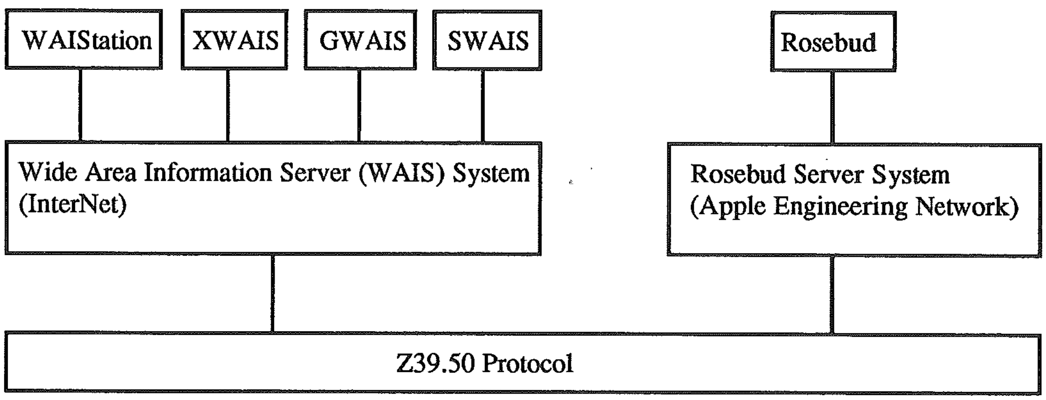


Figure 1 *The interfaces to the WAIS and Rosebud server systems, and the protocol.*

After the collaborative phase of the WAIS project came the Internet experiment. In this phase of WAIS, source code for the open protocol, information servers, and for several interfaces were made freely available over the internet. In addition, Thinking Machines established and maintained a directory of information servers, which WAIS users could query to find out about available information sources. This phase of WAIS is still in progress, and has resulted in the creation of new interfaces, the availability over the internet of more than a hundred servers on three continents, and over 100,000 searches of the directory of servers. In the first 6 months of the Internet experiment, approximately 4000 users from 20 countries have tried this system, with no training other than documentation (Kahle, 1991b). Administrators of popular information servers indicate that they are getting over 50 accesses a day from many countries.

2.3 The WAIS System

WAIS employs a client-server model using a standard protocol (based on Z39.50) to allow users to find and retrieve information from a large number of servers. The client program is the user interface, the server does the indexing and retrieval of documents, and the protocol is used to transmit the queries and responses. Any client which is capable of translating a user's request into the standard protocol can be used in the system. Likewise, any server capable of answering a request encoded in the protocol can be used.

A WAIS server can be located anywhere that one's workstation has access to: on the local machine, on a network, or on the other end of a modem. The user's workstation keeps track of a variety of information about each server. The public information about a server includes how to contact it, a description of the contents, and the access cost.

The WAIS protocol (Davis, 1990) is an extension of the existing Z39.50 standard (NISO, 1988) from NISO. It has been augmented where necessary to incorporate many of the needs of a full-text information retrieval system. To allow future flexibility, the standard does not restrict the query language or the data format of the information to be retrieved. Nonetheless, a query convention has been established for the existing servers and clients. The resulting WAIS Protocol is general enough to be implemented on a variety of communications systems.

The WAIS clients will be described in detail in the next several sections. However, all of them work in a basically similar way. On the client side, queries are expressed as strings of words, often pseudo natural language questions. The client application then packages the query in the WAIS protocol, and transmits it over a network to one or more servers. The servers receive the transmission, translate the received packet into their own query languages, and search for documents satisfying the query. The lists of relevant documents are then encoded in the protocol, and transmitted back to the client. The client decodes the response, and displays the results. The documents can then be retrieved from the server. The documents can be in any format that the client can display such as word processor files or pictures.

3. WAISTATION: AN INTERACTIVE QUERY INTERFACE

WAIStation At A Glance	
Target Machine	Macintosh Plus and above, 9" Monochrome screen.
Effort	1 man-year
Number of Users	2000
Status	finished, freely distributed
Language	ThinkC
Communications	TCP/IP and Modem (not supported)
Designer	Harry Morris
Organization	Thinking Machines
Availability	Available for anonymous FTP from /public/wais/WAIStation*.sit.hqx@think.com
Design goals	Implementable quickly, support interactive queries well, changeable based on user's comments, make something very simple to learn (partner friendly), try out many ideas: interactive queries, passive alerting, asking multiple servers.
Used	In a study with accountants and tax consultants at KPMG: very good user acceptance. In the Internet experiment: estimated that half of the uses of WAIS are using WAIStation. (based on when the directory of servers did not work for Macintoshes, usage dropped to half).
Problems	dealing with the directory of servers (s). Modem code was difficult to get right.

WAIStation was designed for use in the WAIS experiment at KPMG Peat Marwick. As such, we needed an interface that would be easy to use, and would encourage successful searches by users untrained in search techniques. Peat Marwick often sends its employees into the field toting their Macintosh SE's along for use as portable computers. Thus we had to design the interface to run on a 9-inch black-and-white screen, and make minimal demands on CPU and memory. Furthermore, WAIStation was designed for use over modems and slow LANs.

3.1 Design Rationale

In designing WAIStation, we were informed by two metaphors - search as conversation, and storage by file folder. The process of formulating an effective search is highly interactive. Of the documents which match a query, the ones which match "best" are displayed. One or more may be of interest, in which case, they can be fed back to the system, interactively improving the search. We choose to view this process as a conversation. Thus the initial natural language question becomes that starting point for give and take between the user and the server(s). Relevance feedback provides the context for the question. As the search proceeds, some results may suggest alternative searches or branches of the conversation. This is provided for by allowing several questions to evolve at the same time.

Eventually one or more questions may be refined to the point where they are finding consistently good results. At this point, the question can be automated, becoming a dynamically updated file folder. At intervals these questions wake up and query their servers. The results are stored in the results field for later inspection. They can be thought of as regular Macintosh folders, except augmented with a charter describing how to keep their contents up to date.

This parallel with the Macintosh folder structure suggested a drag and drop construction for the user interface itself. Constructing a question is a three step process - typing the key words, specifying the servers to use, and specifying the relevant documents to feed back. If we think of questions like Macintosh folders, we can use the Macintosh's drag and drop mechanism for putting sources and relevant documents into a question. This approach makes WAIStation's mechanics instantly familiar to users of the Macintosh finder.

3.2 Human Interface

When WAIStation starts up, two windows appear – one contains the users available Sources (see below) and one contains the users saved Questions. Sources are identified by an eye icon, questions by a question mark icon.

Double clicking on a question icon opens the stored question, including any new results found since the last time it was examined. The top half of the question window contains a field in which to type key words (the natural language part of the question), a list of relevant documents, a list of sources, and a list of result headlines. Sources can be added to the question buy selecting a source icon (in the Sources window), and dragging it into the question. Relevant documents are specified in the same way.

Result documents, returned by the servers, can be examined by double clicking on their icon. Note that the result list contains a graphical indication of how well each document matches the query. The original graphic was a series of 0 to 4 stars, similar to the ratings found in TV guide. We thought that this rating scheme would be easily recognized. Experience proved that the stars did not provide enough information to be recognized, or to discriminate among the documents. Latter versions of the software replaced the stars with a horizontal bar giving 20 levels of resolution.

Any of the resulting documents can be opened and viewed in its own window. WAIStation supports plain ascii documents as well as PICT format pictures. Text windows automatically scroll to the position which the server considers the most relevant part of the document. This allows the user to quickly determine if a file is useful. In order to perform well over slow communications channels (modems and slow LANs) the text is downloaded on demand in 15 line chunks. The keywords used in the query are automatically highlighted in bold.

Sources are specially formatted text files which describe information servers and how to get to them. Double clicking on a source displays a window with several controls. The top part is information specified by the server itself - a pop-up menu to specify the method of contacting the server (ip-address/tcp-port, modem number and speed, or location of a local index); a script to run after logging in (for use by modems); a database to search (servers can support multiple databases); a display of when the server is updated, how much it costs to search, and a textual description of the databases' contents. The bottom half of the source window allows the user to specify personal information about the server - when to contact it (for automatic update); when it was last contacted; how much to spend on it; how much credence its results should be given (this is used to scale document scores, which helps in the sorting of responses to questions asked of multiple servers); the number of documents to ask for when searching it; and finally the font and type size to use when displaying plain text results (important to publishers). Several of these fields are merely place holders in the current implementation. In particular, budget and confidence have not been implemented yet since there are no for-pay servers yet, and the number of sources is still relatively small.

Source files can also be retrieved from servers. This allows users to search servers whose database elements are pointers to other servers. The results can be used as targets for further searches. An experimental directory of servers is being maintained on the Internet.

3.3 Implementation

WAIStation was implemented in Think C 4.0 using the object oriented class library. It took about a man year of effort. The most difficult parts were the automatic update facility and the communications. Automatic Update required the ability to do background processing - which is not a normal part of the Macintosh operating system. Communications were difficult primarily because we were simultaneously debugging the Z39.50 protocol, modem code, and the (then new) Apple Communications Toolbox. We eventually left modems unsupported, and replaced the Communications Toolbox with direct calls to MacTCP. Through this experience we found that communications speeds of less than 9600 baud were barely tolerable for interactive text retrieval.

3.4 Observations

We estimate that WAIStation is now in use by over 2000 users in twenty countries. The common user complaints center around configuring MacTCP, using (the undocumented) directory-of-servers, and avoiding a bug requiring the software to be installed on the start up disk.

We have noticed several shortcomings in the current design:

- Users want access to their own data - WAIStation is capable of searching a Macintosh based inverted index file, but we unbundled the index builder when we realized how much work it would take to make it useful under Macintosh OS. OnLocation (On Technology) is an implementation of a Macintosh indexer that could be used.
- Interaction with the directory of servers is incomplete - It is not obvious which search results are source files, and what to do with the ones that are. It should be possible to drag a retrieved source directly into a question's source window, but the present interface requires that it be saved first. The lesson we learned was that special cases should be handled specially, rather than forcing users to use general techniques "for consistency's sake".
- Printing documents and searching for keywords in documents (find/find-next) are simple functions which users expect.
- People want to see their documents in their original form - WAIStation currently only displays ascii and PICT. This can be fixed with format filters such as Claris' XTND, at the expense of the ability to download arbitrary sections of a document, since such filters require that the document be processed from the beginning.
- Relevance feedback was not obvious - users unfamiliar with the use of relevance feedback did not think to use it - it needs to be made more automatic. One way to do this might be to extend the notion that a question is a conversation, with relevance feedback as context (or body language) - clients or servers can be written that watch their users, and deduce which documents were relevant based on which ones were read. A simpler approach might be to always do relevance feedback, presenting the results in a "see also" list. We tried this, but the Macintosh was too slow to make it useful.
- Communications over 2400 baud modems are too slow to support interactive queries. We found that 9600 baud is barely acceptable, while 56Kb is sufficient to support several users.
- The finder-like interface (drag and drop) is not obvious - Even though the Macintosh Finder is based on drag and drop, no one expected it in an application. Once users were shown what to do, it was very natural. It was also not necessarily the best use of screen space, since it required that both the start and end of the drag be visible on the screen at the same time. Another anomaly worth mentioning is the fact that although we were simulating the finder, we had no "trash can" analogy. Removing a source was accomplished by dragging it onto the desk top and dropping it there, which confused some users.
- The alerting system was crude. For example, there was no visual cue to tell the user that a question had found new documents in the background. Also, the background searches did not exclude previously read documents.
- Headlines often don't give enough context - The headlines displayed in the question window were only about 60 characters long, making it difficult to identify which documents were useful without opening them. Furthermore, there was no provision to display the document's date or the name of the source it came from.

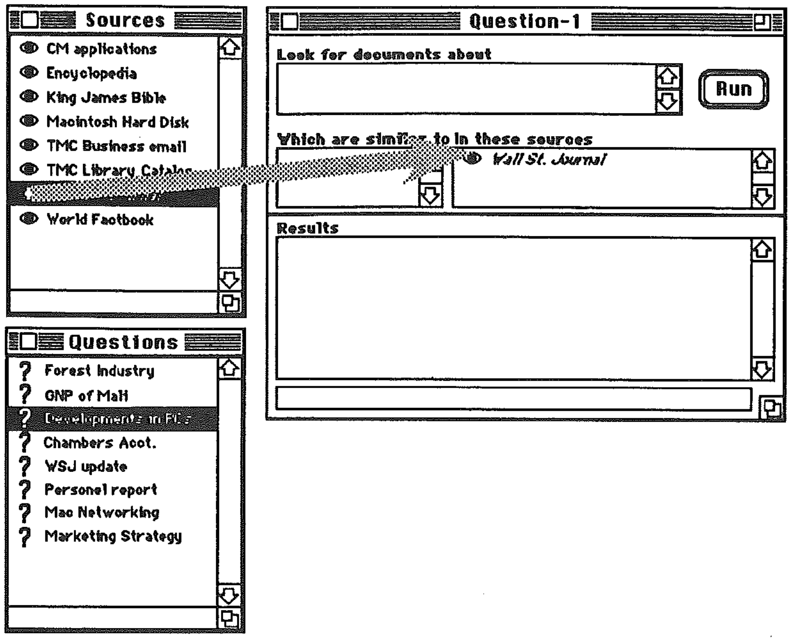


Figure 2 *WAIStation's Sources and Questions windows store the user's personal objects. Dragging a source into a question window specifies that the question will contact the source in order to fulfill its charter.*

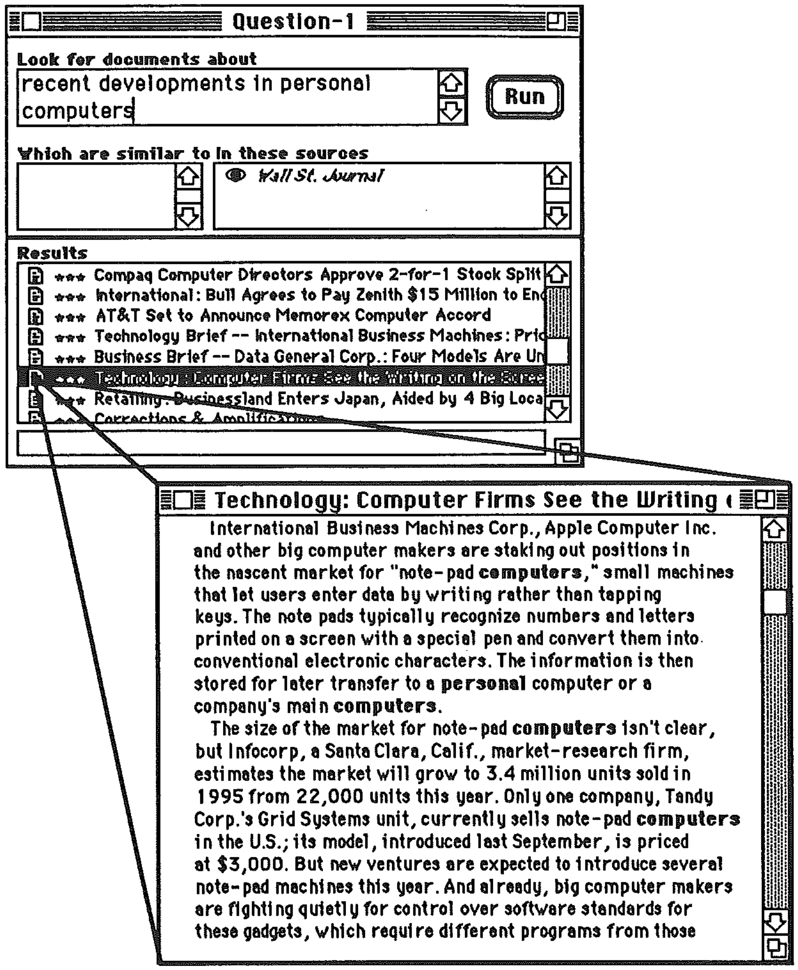


Figure 3 *After running the question, results are displayed in a scrolling list. Double clicking on a result opens a document window. Query words are highlighted.*

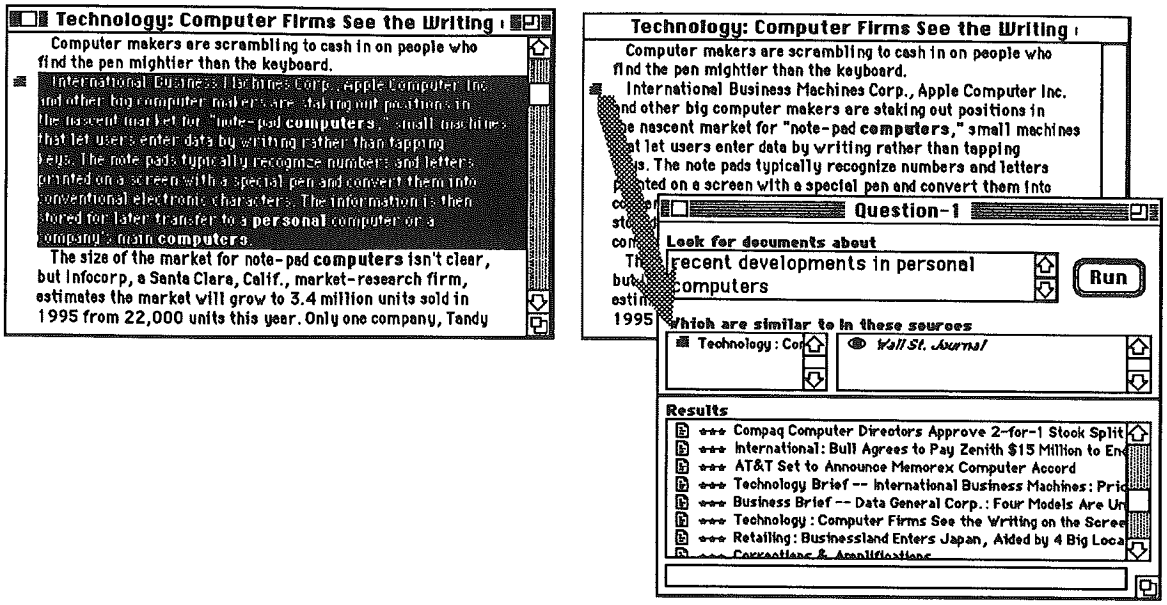


Figure 4 *Relevance feedback is done by selecting a document or part of a document, and dragging the document or paragraph icon into a question.*

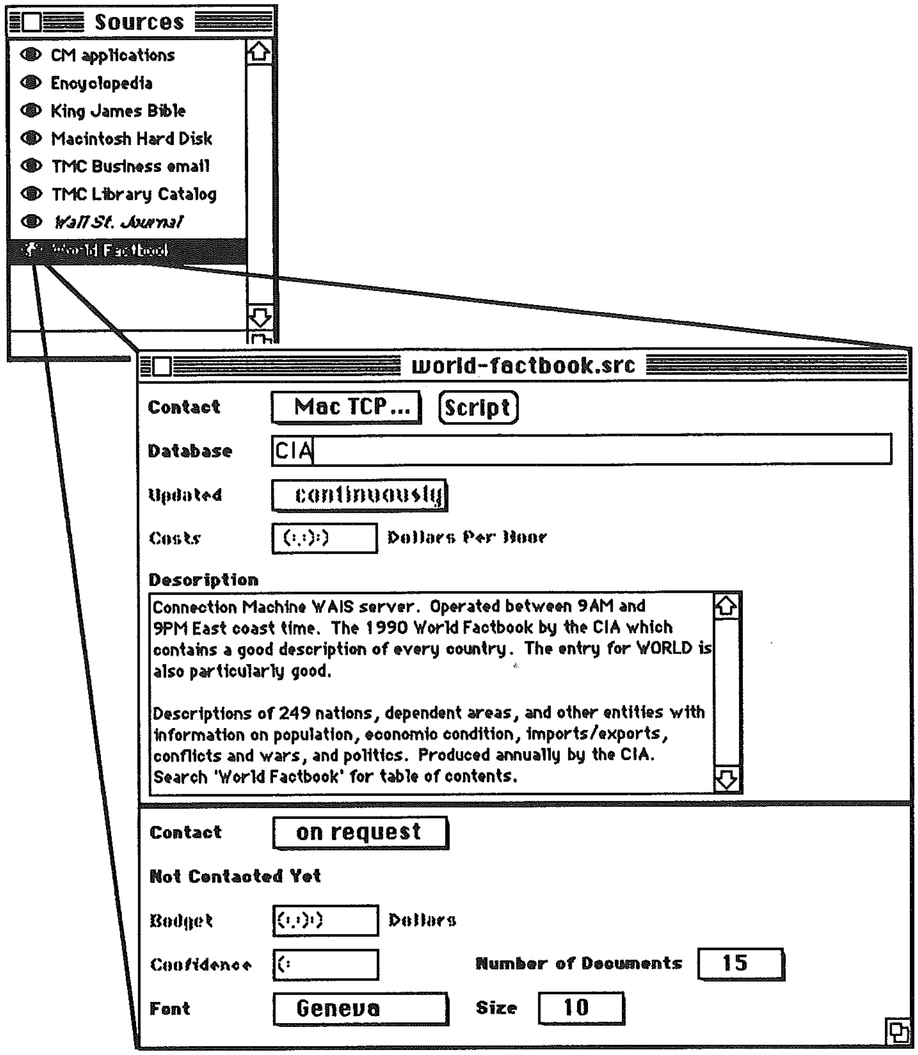


Figure 5 Double clicking on a source icon opens a source window.

4. X WINDOWS BASED INTERFACE FOR WAIS: XWAIS

XWAIS At A Glance	
Target Machine	X-windows terminals on unix machines
Effort	4 man-months
Number of Users	500
Status	finished, freely distributed
Language	C
Communications	TCP/IP
Designer	Jonathan Goldman
Organization	Thinking Machines
Availability	Available anonymous FTP from /public/wais/wais*.tar.Z@think.com
Design goals	Copy WAIStation so that we can leverage one design, portable and based only on freeware Display data in many different formats (image, text, etc)
Used	Used in the internet experiment Heavy use by X users within Thinking Machines and outside
Problems	Installing it has caused many users to stumble. The number of variables (architectures, X directory structures) makes it difficult to make it portable touch on the ability to handle different types (this is unique to this interface). uses other programs to help (like interapplication communication)

The WAIS interface for the X Windows environment was developed for the Internet experiment to provide an X Windows based interface for a growing community. It was built to look as much like the Macintosh WAIS interface (WAIStation) as possible, given the limitations of the freely distributed X Windows software. Since the metaphors in XWAIS are nearly the same as those for WAIStation, a user of one system can easily move to the other, without having to learn much new. In fact, the underlying data structures are identical to those in WAIStation, so questions can be copied from a Macintosh to a UNIX machine running XWAIS, and used without modification.

XWAIS supports interactive WAIS access, including question entering, source selection, addition of relevant documents and pieces of documents. Unlike WAIStation, XWAIS retrieves an entire document when requested, instead of just the parts being viewed. We decided this was acceptable, since the underlying networks for X will most likely be fast.

Since XWAIS runs under X windows, and was built for the UNIX operating system, it can take advantage of the tools available for these systems to display a wide range of document formats. A simple filter interface is provided in the application (as an X resource) to allow a user to select the tool required for a given type of document, e.g, if the document is a postscript file, xps can be used to view it. This is a feature that is not available in any of the other user interfaces described here.

In order to distribute this software without restriction, XWAIS uses the freely distributed Athena Widget set included in the X11R4 release from MIT. Although these widgets don't look as nice as some others that are available, they can be used to build a useful interface. Some aspects of this interface are restricted by the nature of the widgets available. XWAIS was built using the Xt X Toolkit Intrinsics, and allows a large amount of customization of the appearance of the display using X resources. The application relies heavily on the Xt resource mechanism, and will not run unless these resources are in place. The "object-oriented" feel of these widgets made building the interface rather easy, once the widget with the closest desired functionality was found. Finding the correct widget was the hardest part. Most of the actual behavior of the interface is controlled by "call-backs" - the methods that widgets inherit.

The XWAIS application is actually two separate applications: XWAIS, a simple shell for selecting sources and questions, and xwaisq, the application that actually performs WAIS transactions. The C code in xwaisq is also used

in waisq, the shell-support program for GNU Emacs WAIS. This allows users to use simple UNIX facilities to submit questions created by xwaisq using waisq (e.g. a crontab entry to periodically query a server).

The implementation for XWAIS was done in C (6k lines), using the X11R4 release of X windows from MIT, the Xt X Toolkit Intrinsics, and the Athena Widget Set, included in the X Windows release.

XWAIS is a text-based user interface built in a graphical window environment. Some additional graphical metaphors would be desirable, but the limited widget sets precluded that. It would take a considerably larger amount of work to add much graphics to this application. Perhaps some other X toolkit would provide simpler methods for doing this.

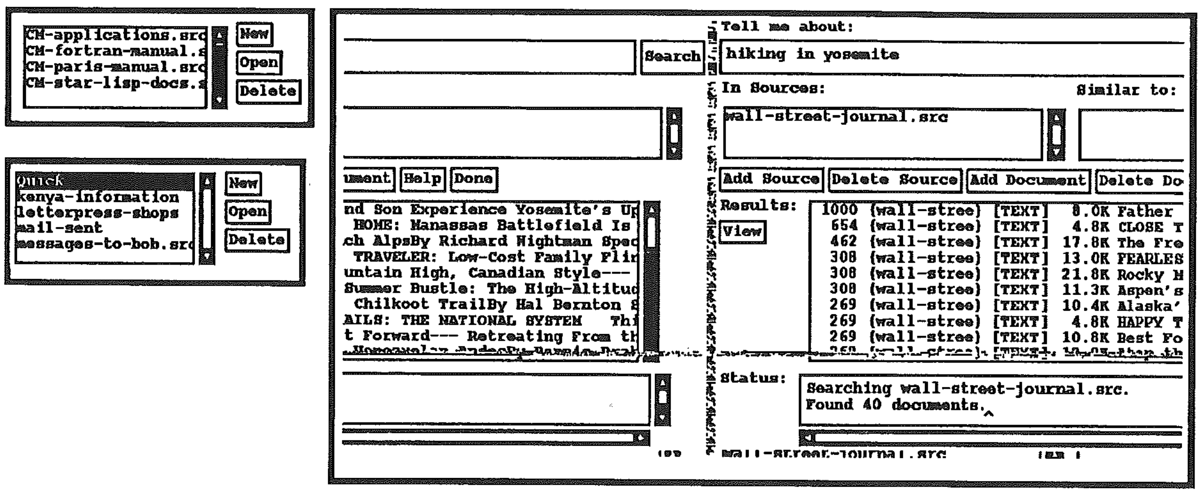


Figure 6 *The XWAIS interface, including the Questions and Sources windows, and an open question.*

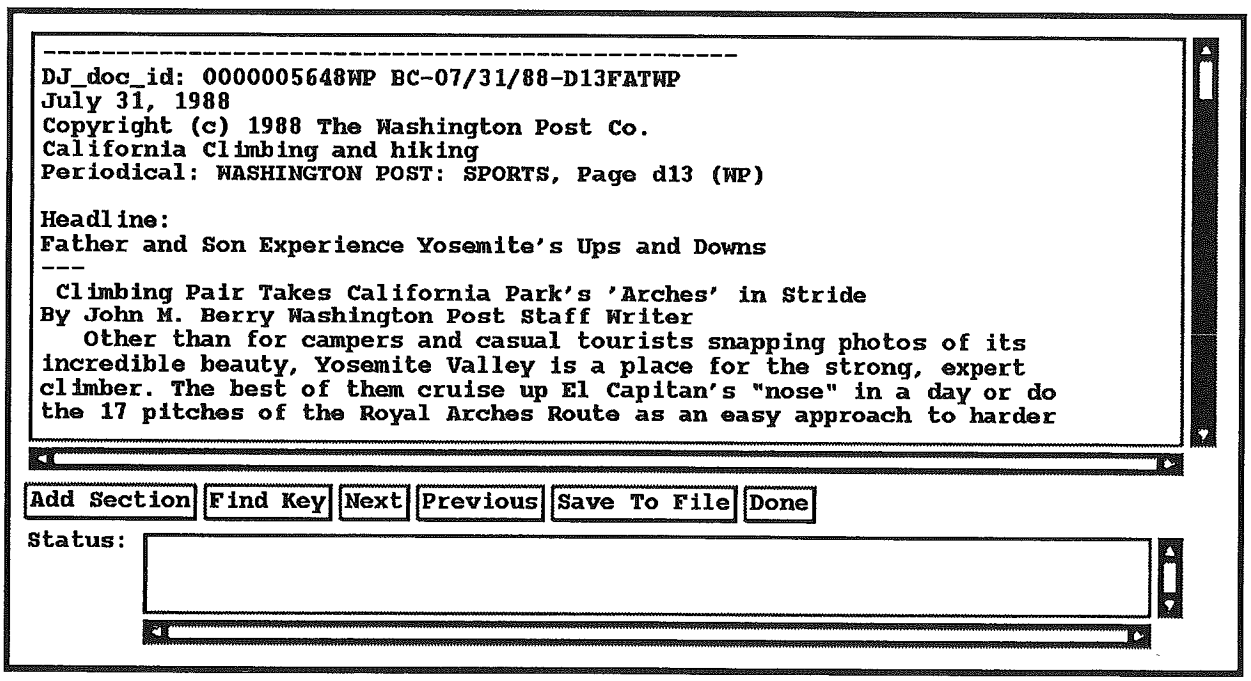


Figure 7 *A document displayed in the XWAIS interface.*

5. GNU EMACS WAIS INTERFACE: GWAIS

GWAIS At A Glance	
Target Machine	terminals on unix machines
Effort	2 man-months
Number of Users	500
Status	finished, freely distributed
Language	gnu-lisp, and C
Communications	TCP/IP
Designer	Jonathan Goldman
Organization	Thinking Machines
Availability	Available anonymous FTP from /public/wais/wais*.tar.Z@think.com
Design goals	Copy WAIStation so that we can leverage one design, Use precedent from other gnu-emacs applications: RMAIL, dired
Used	Used in the Internet experiment with heavy use by some gnu-emacs users
Problems	Dealing with the directory of servers. Using passive alerting

The WAIS interface on GNU-Emacs/Unix (GNU) was developed specifically for the Internet experiment for a technically strong user population. The reasons it was developed were: the large number of emacs users, the extensibility, the ubiquitous nature of character display terminals, and the component nature of emacs which meant WAIS could be integrated into email, bboards, and programming tools.

The design of the interface was a cross between WAIStation and other emacs interfaces. The direct manipulation of WAIStation was replaced by command keys, as is common in emacs applications. The choice of command keys were modeled on the dired and RMAIL emacs applications.

GWAIS allows users to access the interactive features of WAIS: question entering, relevance feedback, displaying document, and source selection. An extra feature, not found in the other interfaces, is an interface to an indexer for creating sources, but it appears that this feature is not heavily used. Furthermore it allows questions to be saved, but it depends on the user to automate the update of questions and sources using cron or other Unix tools. Graphic documents can be displayed on X Windows terminals if the user has set up the environment variables.

The implementation of GWAIS was in emacs lisp (2K lines) and in C code (3K lines). About half of the time of a typical search and retrieval is spent in reading the data into lisp.

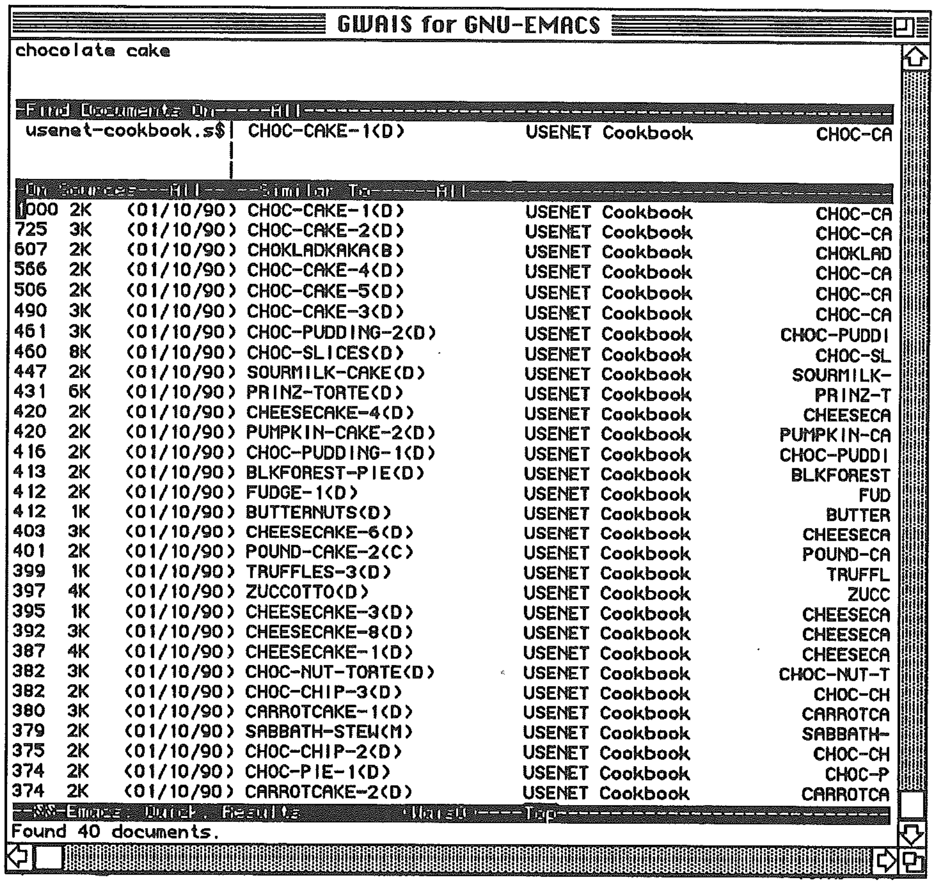


Figure 8 *The GWAIS interface, displaying the results of a relevance feedback search.*

6. SCREEN BASED (TERMINAL) WAIS INTERFACE: SWAIS

SWAIS At A Glance	
Target Machine	Terminals connected to Unix systems
Effort	1 man-month
Number of Users	900
Status	beta
Language	C
Communications	TCP/IP
Designer	John Curran
Organization	NSF Network Service Center
Availability	To be included in WAIS release, anonymous FTP from /public/wais/wais*.tar.Z@think.com
Design goals	Highly Portable, Provide straight-forward user interface, Utilize existing application key mappings (rn, vi, emacs), Support multiple servers per query, Allow for personal "source" directory and a common source directory, Allow for useful source discovery via searches, Provide simple active tool with little state (no question storage, relevance feedback, or passive notification)
Used	Internet users via telnet: k-12 students, educators, user services staff, librarians, and (occasionally) network staff
Problems	Dealing with the directory of servers. Lack of information in many server-returned records. Providing simple and uniform nomenclature Planning for large numbers of sources.

To open WAIS to a wider community of users, an interface was developed to run on dumb terminals or over telnet sessions. It is called "SWAIS" for Screen WAIS since it is uses a character display terminal screen for the interface. The user communities that this interface can serve are dial-in users, telnet users, and low-end terminal users.

The design of the interface involved 3 screens: a single screen listing all known servers that the user could pick from; a list of search result documents headlines; and a document display screen. Listing all servers and allowing users to pick which servers to use encourages users to ask questions of multiple servers. Unlike the other interfaces, the sources list shows what site runs it and how much it costs (if anything). The resulting document screen includes headlines and how many lines it is, but its innovation is to show what source it came from.

It does not handle relevance feedback or downloading new sources from the directory of servers. Another drawback is using it with large numbers of sources since moving around the list requires scrolling. On the other hand, this server has proven to be very popular on the Internet because of its ease of use, all a user has to do is telnet to a specific machine to use it.

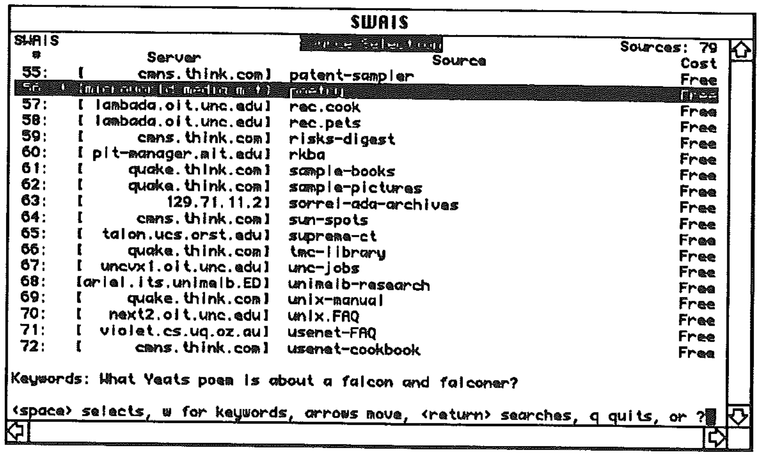


Figure 9 *The SWAIS query building screen. The poetry source is selected, and search terms are entered. This interface does not currently support relevance feedback.*

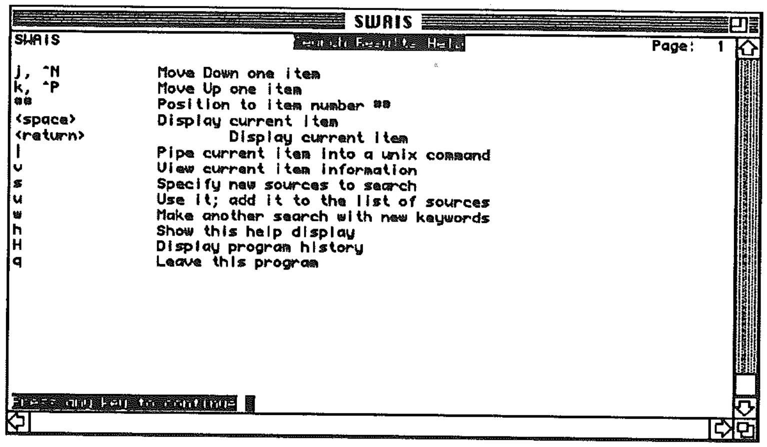


Figure 10 *The SWAIS help screen.*

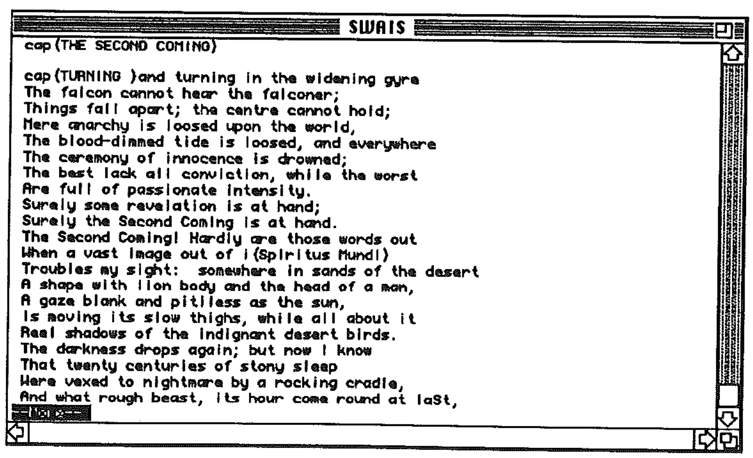


Figure 11 *A document displayed in SWAIS.*

7. THE ROSEBUD INTERFACE: REPORTERS AND NEWSPAPERS ON THE MACINTOSH

Rosebud At A Glance	
Target Machine	Macintosh II, color screen
Number of Users	25
Status	Finished; internal use
Language	Smalltalk, MPW-C
Communications	TCP/IP using IPC package
Designers	Charlie Bedard, David Casseres, Steve Cisler, Tom Erickson, Ruth Ritter, Eric Roth, Gitta Salomon, Kevin Tiene, Janet Vratny-Watts.
Organization	Apple Computer
Availability	Only internally to Apple ATG
Design goals	Serve as research platform for interface and architectural explorations. Allow ordinary users to create personalized information flows; support passive alerting, scanning and capture of information.
Used	Used in various internal tests; not available for the Internet experiment.
Problems	No good interface mechanisms for providing users with convenient access to large numbers of servers.

Rosebud is a project within Apple Computer's Advanced Technology Group. Its principle objective is to serve as a platform for investigations into what is needed to make remote information accessible and useful to ordinary Macintosh users. The investigations have two foci: human interface components and techniques; and system architecture issues. In this article we focus exclusively on the human interface aspects of Rosebud.

The Rosebud Server System is similar to the WAIS system in that it uses the Z39.50 protocol to access multiple, remote database; it differs from them in that it contains extra underpinnings for making information access an integral part of the Macintosh environment. Specifically, the Rosebud Server System allows users to create autonomous, ongoing "agent" processes which access, update, and present information from local and remote sources. The Rosebud system does not currently provide access to the internet WAIS servers (for reasons of network security, rather than basic incompatibilities), and is not publicly available.

7.1 Design Rationale

The design of the Rosebud interface began with a study of the practices and problems of ordinary information users. The principle focus was on information users at KPMG Peat Marwick in San Jose, the original client site for WAIS; in addition, several groups of users of on-line information services within Apple were also studied (Erickson, 1991). Interviews with accountants at Peat Marwick enabled the designers to put together a schematic of how information (mostly paper-based information) flowed through their offices (figure 12).

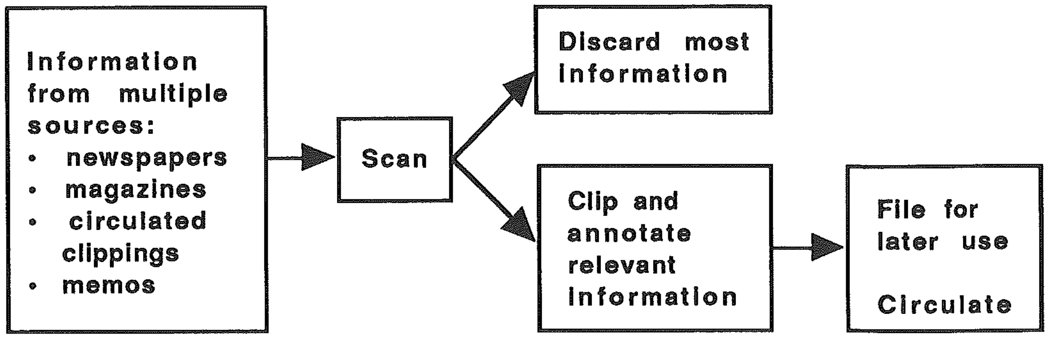


Figure 12 Information flow through accountants' offices.

Several features of this schematic informed the design of Rosebud. First, information typically came to the accountants via newspapers, magazines, and memos; instances where the accountants went out of their way to search for information were less frequent. Second, the accountants never talked about "reading" information; they always spoke of scanning, or skimming it—they didn't have time to read it. This suggested that a good interface should provide a way for the users to scan retrieved information quickly. Third, accountants remarked that they discarded most information, including information that might be useful. Potentially useful information was discarded for two reasons: the accountants didn't have the physical space to store everything, and they knew from experience that if they tried to save too much, they wouldn't be able to find anything later, when they actually needed it. This suggested that giving users access to remote information was just half the problem; users also needed tools for archiving, organizing, and re-retrieving information. Finally, when users did come across information that seemed worth saving, they would typically cut it out (the accountants used, almost exclusively, paper-based information), and then they would annotate it by circling, underlining, or jotting a few notes in the margin. Annotation turned out to be an important concept: not only did it help the user who annotated when the information was re-retrieved later on, but it also helped others scan the information more quickly when copies were passed on to them.

The consequence of these observations was a design for a system which allowed users to define topics of interest which would be automatically retrieved, and would then permit them to scan those items and save them into an environment where they could be annotated, organized and re-retrieved.

7.2 Human Interface

The Rosebud interface design has three components: reporters, newspapers, and notebooks. Reporters are for retrieving information. Users give reporters assignments which specify what to look for, and where to look. This is shown in figure 13: users enter words describing the information in which they're interested, check off the information sources they wish the reporter to search, and, if they so choose, automate the reporter so that it searches the databases on a daily or weekly basis. Upon pressing the "Search" button in the assignment window, a reporter is created, performs the search, and returns with a list of results (figure 14).

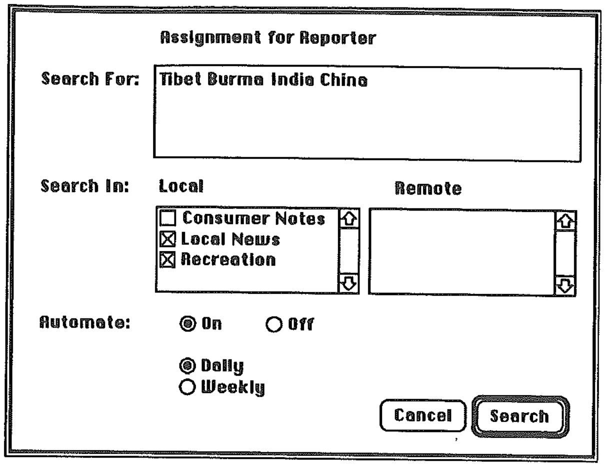


Figure 13 Creating a reporter—the assignment window.

The reporter window (figure 14) provides users with a variety of ways to look over their results, and refine their queries. The results are shown in the "Best Guesses" pane. (The name '"Best Guesses" was chosen to provide some indication that inaccuracy could be expected; our observations of users had shown that they were often mystified by some of the items that showed up as the results of searches.) The asterisks to the left of items indicate their relative relevance, and the pop up menu above the pane allows users to order the list by date or relevance. Simply selecting an item shows a preview of it—a short excerpt with search terms highlighted in color and boldface (figure 15). Previews are useful because users can get a look at a little bit of the item without incurring the overhead of downloading the whole article over the network. Users also have the options of saving articles to their disks or opening them for viewing. Finally, having looked over their results, users can refine their search in the bottom pane of the window.

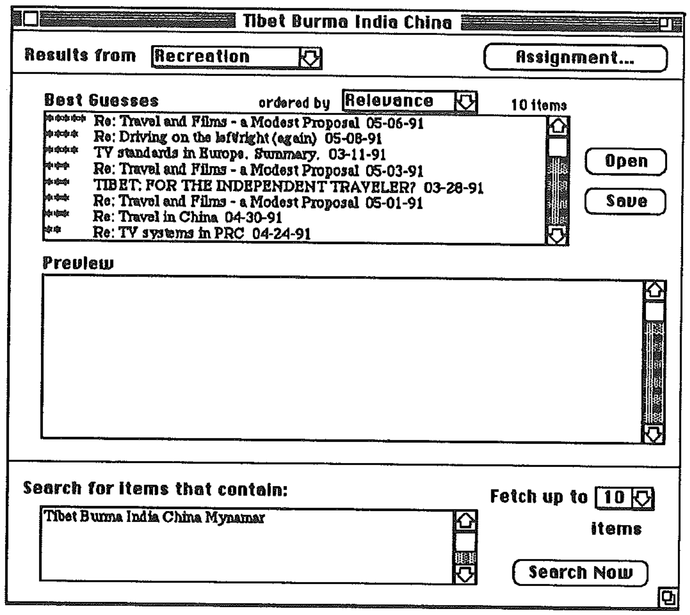


Figure 14 The reporter window contains the results of the search and provides means for previewing, opening, and saving results.

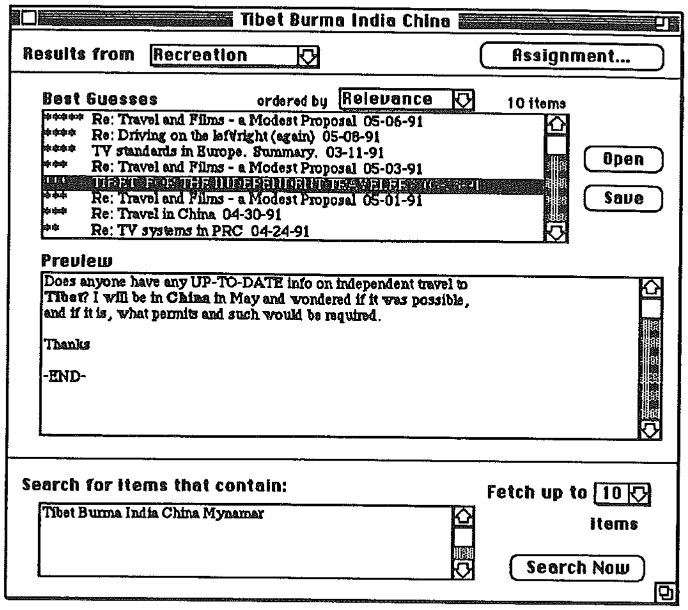


Figure 15 *The reporter window makes it easy to scan through hits. Clicking on a retrieved items generates a preview which shows an excerpt to the hit with the search terms (Tibet and China) highlighted in boldface. The user can refine the query in the lower pane of the window.*

The above sequence occurs whenever a user creates a new reporter. However, since users are likely to use many reporters, and because the initial user studies indicated that ways of skimming through incoming information were important to the accountants, the newspaper was provided to support rapid scanning of new information. The model of a newspaper is quite simple (figure 16): on the left is an index column which contains the names of all reporters, and to the right are two columns of news. Each reporter 'owns' one news column and publishes the title, date and an excerpt of each item in its column. The columns scroll independently, using 'minimalist' scroll bars to prevent the multiple scroll bars from visually overloading the screen. If an excerpt seems interesting, double clicking on it opens the full article in a window, from which it can be viewed, printed, or saved. Thus, rather than having to open up a dozen reporters every morning to see what's new, the user can go to one place, the newspaper.

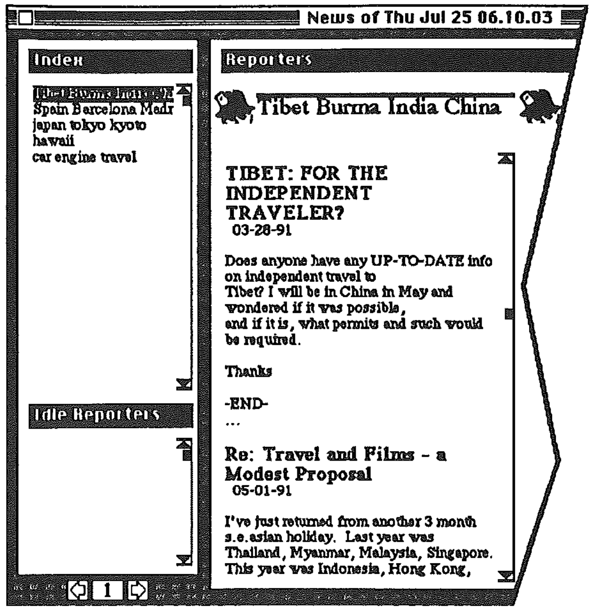


Figure 16 *The newspaper allows users to quickly scan through new items retrieved by the reporters which are working automatically.*

The newspaper can also serve as a control center for the Rosebud interface. The user can open a reporter by clicking on its name or icon at the top of its news column. Consequently, if a reporter's column has strayed from the desired topic, the user can quickly get to the reporter and revise its assignment. The index also lists inactive reporters (those either not automated, or that haven't found anything new since the last newspaper), so they too can be opened, and automated or otherwise adjusted.

A third component of the Rosebud interface—the notebook—was designed but not implemented. Notebooks are environments within which users may save, annotate, and organize retrieved information. Notebooks were designed in response to the observations of Peat Marwick accountants, which indicated the need for an environment which supported the way accountants worked—in particular, notebooks were intended to support annotation, and re-finding retrieved information at a later date. A particularly nice feature of the notebook design was its use of annotations as landmarks for re-finding information. The notebook design, and its rationale, is described in (Erickson, 1991).

7.3 Implementation

The Rosebud system consists of six parts: 1) a human interface application written in SmallTalk/V (to facilitate the rapid changes in the interface necessary to effectively conduct interface design research); 2) a search manager package which implements the autonomous agent functionality and formulates Z39.50 queries for 3) remote Z39.50 servers implemented in MPW C that automatically index items placed in their input folders by 4) HyperCard stacks that download new items from a Net News server; 5) a file manager component (MPW C) that does all of the file I/O and compaction for reporters and newspapers; and 6) directory servers which allow the various components to find one another. All of these components are written as separate applications and communicate with one another using a prototype IPC that runs over TCP/IP. The file manager and search manager applications run in the background under MultiFinder, enabling Rosebud to access information and construct newspapers while the human interface application is not running. Like the other WAIS interfaces, Rosebud uses the WAIS protocol package. The human interface was designed for Macintosh II class machines, with 13 inch color screens.

7.4 Observations and Testing Results

The Rosebud human interface was subjected to informal testing on 14 users. Users were told only that Rosebud was an application for finding information, and then given a particular topic to find information on. They were given no help or documentation. Note that although informal, this type of testing is very stringent, in that users

approach the application knowing almost nothing about what it is, or why they would actually use it. Data collection consisted simply of recording their questions, observations, and problems as they went along, administering a post-test questionnaire, and then asking them a few, open-ended questions. Here are a few of the more general observations.

- Over 80% of those who tried the Rosebud interface responded very positively to it, and said that they would use something with its capacities as part of their daily work routine. Two thirds of users indicated that they would usually use newspapers to browse through information (instead of reporters).
- At the end of the test, over two thirds of the users said they liked the metaphors of reporters and newspapers; however, almost all users had some difficulty in getting started. The typical problem was that users did not associate reporters with a way of retrieving information. When asked to find information, users first looked for an item called search; when they didn't find this, they usually turned to the newspaper, which is, in fact, where they look for information on a daily basis. It is possible that this problem can be remedied by minor interface changes (e.g. putting a "New Reporter" item in a search menu); alternatively, it may be that the metaphor is inappropriate.
- A number of users were lead astray because they had conceptual models of information retrieval based on their familiarity with query languages and structured databases. Such users tended to be wary of entering search terms because they weren't sure of what the appropriate syntax was, and didn't understand what "relevance" meant. Those that did know what relevance was wanted to know how the information server calculated it.
- Users liked previews a lot—especially the feature of highlighting keywords in boldface. They wanted to see boldface keywords in the newspaper and article windows. Users also wanted the ability to select text in the newspaper and article windows and change the style or font themselves, so that they could annotate significant items. This parallels practices observed in our initial observations of accountants, where we found that annotation plays several important roles.
- A variety of low level interface problems, due to terminology or graphic design were discovered. Some examples: users did not usually recognize the asterisks in the "Best Guesses" window as indicators of relevance; users didn't think that "idle reporters" was a good name, and said that it was very important to distinguish between reporters which had found nothing, and those which weren't looking.

The testing described above focused on how usable Rosebud was when users were first exposed to it. In the next phase of testing, a small set of users will be observed over the course of a month, in which they have the option of using Rosebud from their desktop machines to access meaningful data. This phase of testing will allow a more realistic assessment of Rosebud, in that it will last long enough to permit users to build up their own set of reporters, and to access newspapers which contain information of personal import.

8. CONCLUSION

This paper has described five interfaces developed to provide access to distributed systems of information servers. The interfaces presented here were developed with different constraints in mind, so it is not useful to compare them directly; instead they may serve as examples of differing responses to issues such as screen size, workstation power and intelligence, communication speeds, and user needs and practices.

The interfaces designed so far have addressed some of the critical issues for end-users to accomplish interactive searches in a wide area network. These include ways of finding which information servers contain relevant information, supporting searching by ordinary users, and supporting browsing of, and passive alerting about, newly retrieved information. The alerting aspects of the interfaces have not been tested much in this environment due to the lack of appropriate data sources for this type of searching. It is probably fair to say that any of the design solutions described here can be improved upon by further work.

The WAIS Internet experiment has revealed a number of issues requiring further work. In the Internet environment we have observed (in the logs of user queries) that users have a difficult time finding out what is in a database, thus demonstrating that there is a lack of browsing or scanning facilities in the interfaces, protocol, and servers, as well as a general shortage of descriptive information about databases.

Finally, there are a variety of other issues raised during the studies of the Peat Marwick accountants which have received little or no work. Document layout is one such problem. Accountants mentioned that sometimes they want to retrieve documents not because of the information they contain, but to look at their layouts (accountants will often examine successful proposals to a client when preparing a new proposal). More generally, users regard pictures, diagrams, tables, and charts as essential components of a document's content. Unfortunately, support for different document formats, and for the retrieval and display of non-textual information within them is very limited on most existing clients.

Another issue is called the boilerplate problem. Accounting documents often contain a large amount of boilerplate, standard text which varies little from document to document. What tools are needed to allow users to effectively retrieve, order, and browse a large set of documents which are 95% similar? Note that boilerplate is characteristic of a wide variety of business proposals and legal documents, not just accounting documents. In fact, the analog to boilerplate occurs in scientific documents in which standard terms and descriptions are used to describe procedures and methods used in an investigation.

A number of other issues remain to be addressed. Users are very interested in being able to see what queries other users are conducting, and what information servers and articles are most popular. A frequent suggestion is to allow users to rate the 'goodness' of articles they retrieve. However, in a commercial setting, information about the kind of questions being posed by a particular company or person can be revealing and valuable. Clearly, the utility that such information could provide must be balanced by concerns about confidentiality and privacy, and mechanisms for user control of descriptive information are essential. Other issues include how to control the pricing, copyright, and distribution issues which accompany 'for-pay' information.

In summary, there is an immense amount of work to be done. A central part of this work involves further research and development of interfaces. We have made the WAIS system publically available in the hope that designers will find that it—with its common protocol and defined infrastructure—can serve as a platform from which to pursue these, and other, research issues.

FOR MORE INFORMATION ON THE WAIS SYSTEM

The success of a distributed system of information servers depends on a critical mass of users and information services. In order to encourage development and use, Thinking Machines is making the source code for a WAIS protocol implementation freely available. While this software is available at no cost, it comes with no support. We hope that it will facilitate others in developing servers and clients.

For more information, please contact:

Barbara Lincoln (barbara@think.com)

Thinking Machines Corporation
1010 El Camino Real, Suite 310 245 First Street

Menlo Park, CA 94025
415-329-9300

Cambridge, MA 02142
617-234-1000

REFERENCES

(Davis, 1990) F. Davis, B. Kahle, H. Morris, J. Salem, T. Shen, R. Wang, J. Sui, and M. Grinbaum, "WAIS Interface Protocol Functional Specification," Thinking Machines Corporation, April 1990. Available via anonymous ftp: /pub/wais/doc/wais-concepts.txt@quake.think.com or wais server wais-docs.src

(Dongarra, 1987) J. Dongarra and E. Grosse, "Distribution of Mathematical Software Via Electronic Mail," CACM, 1987, Vol. 30, pp. 403-407.

(Egan, 1989) D. Egan, J. Remde, L. Gomez, T. Landauer, J. Eberhardt, and C. Lochbaum, "Formative Design-Evaluation of SuperBook," ACM Transactions on Office Information Systems, January 1989.

(Erickson, 1991) T. Erickson and G. Salomon, "Designing a Desktop Information System: Observations and Issues," Proceedings of the ACM Human Computer Interaction Conference, 1991. ACM Press, April 1991, pp. 49-54.

(Ginther-Webster, 1990) K. Ginther-Webster, "Project Mercury," AI Magazine, 1990, pp. 25-26.

(GNU) For information on the Free Software Foundation and the GNU project, see: /pub/gnu/*@prep.ai.mit.edu. For emacs see: /pub/gnu/emacs-*.tar.Z@prep.ai.mit.edu.

(Kahle, 1991a) B. Kahle and A. Medlar, "An Information System for Corporate Users: Wide Area Information Servers," April 1991. Online Magazine, September 1991.

(Kahle, 1991b) B. Kahle, J. Goldman, H. Morris, T. Shen, "Electronic Publishing Experiment on the Internet - Wide Area Information Servers" In preparation.

(Malone, 1986) T. Malone, K. Grant, and F. Turback, "The Information Lens: An Intelligent System for Information Sharing in Organizations; In Human Factors In Computing Systems," CHI'86 Conference Proceedings, Boston, MA; ACM, New York, 1986, pp. 1-8.

(Mosis) For more information on the Mosis fabrication facility write to mosis@mosis.edu.

(NeXT) NeXT Computer Inc., 900 Chesapeake, Redwood City, California, 94063, (415) 366-0900.

(NISO, 88) "Z39.50-1988: Information Retrieval Service Definition and Protocol Specification for Library Applications," National Information Standards Organization (Z39), P.O. Box 1056, Bethesda, MD 20817. (301) 975-2814. Available from Document Center, Belmont, CA. Telephone (415) 591-7600.

(ON Technology) ON Technology Inc., 155 Second Street, Cambridge, Massachusetts, 02141, (617) 876-0900.

(VERITY) Verity Inc., 1550 Plymouth Street, Mountain View, California, 94043, 415) 960-7600.

ACKNOWLEDGMENTS

At Thinking Machines: Ottavia Bassetti, Franklin Davis, Patrick Bray, Danny Hillis, Rob Jones, Barbara Lincoln, Gordon Linoff, Chris Madsen, Gary Rancourt, Sandy Raymond, Tracy Shen, Craig Stanfill, Steve Schwartz, Robert Thau, Ephraim Vishniac, David Waltz, and Uri Wilensky.

At the National Science Foundation: Suzi, Alex, et. al.

OVERCOMING THE TYRANNY OF DISTANCE:

ELECTRONIC INFORMATION ACCESS IN AUSTRALIA

Colin Steele
University Librarian
Australia National University
GPO Box 4, CANBERRA ACT 2601
AUSTRALIA
Telephone: (06) 249 2003
Facsimile: (06) 249-0734
Internet: C.Steele@library.anu.edu.au

ABSTRACT

A brief overview is provided of recent telecommunications and networking developments in Australia which have not only provided increased communication and cooperation opportunities for Australian libraries, but also effective linkages to international information sources. The consequences for an effective "distributed national collection" within Australia is examined in this context, as well as academic interaction in the change from storage to access philosophies. Other developments, such as international satellite television provision and electronic short loan facilities, are highlighted.

Background

Australia in its early European settlement period of history suffered from what one of its leading historians, Emeritus Professor Geoffrey Blainey, has termed the "tyranny of distance". Australia's distance from the major Western intellectual and political centres, particularly in the 19th and early 20th centuries, made its inhabitants culturally self conscious, albeit externally self reliant. As a result the phrase, which is still in local currency, "cultural cringe" came into being. Thus speakers from overseas even today can be regarded and feted as "gurus", even when the message that they are propagating is exactly the same as that which local speakers have given. Thus an author such as Peter Carey or Thomas Keneally is regarded with more favour if major reviews come from the New York Times Book Review of The Times Literary Supplement than from the Australian Book Review or The Australian.

Part of the reason for a lack of angst in the information and library profession is that the tyranny of distance has been significantly, if not totally, eroded by the electronic and communications revolutions, for example, e-mail, network connectivity, and satellite television. The 24 hour news service from CNN is only one of the satellite TV services in the Australian National University (ANU) Library. This Library also has live direct satellite information from Russia, France, Indonesia, and China, so that scholars wishing to be kept in touch with changing world events can monitor, and, where appropriate and legal, record for historical purposes the data contained in these services.[1]

More generally, the interconnectivity of local and international networks has brought about a revolution in scholarly and intellectual communication. The development of INTERNET and JANET have been well documented and links to the Australian counterpart AARNet (Australian Academic and Research Network) are extremely important. Taking this into conjunction with the ever increasing power of microcomputer workstations and local area networks the potential for document supply and intellectual communication are obvious as a myriad of articles and communications in the recent years have evidenced.[2]

AARNet Australian Academic and Research Network)

The Australian Academic and Research Network (AARNet) is managed by the Australian Vice-Chancellor's Committee (AVCC), which levied individual universities to establish the network that now links not only the universities, but a number of other important institutions such as the Commonwealth Scientific and Industrial Research Organization (CSIRO) and the National Library of Australia. AARNet has been in operation now for roughly two years and is now part of a "glorious global anarchy".[3] It now means that individual data is transmitted extremely quickly between universities, from the ANU to the University of Melbourne 14 milliseconds; to Brisbane 33 milliseconds; to Perth, at the other side of the continent, somewhere between 87 and 300 milliseconds. To reach San Francisco is about two-thirds of a second and Oslo is connected in 957 milliseconds.

The rate of network traffic growth has been, as elsewhere, extremely rapid. Huston, the Technical Manager of AARNet, reported at the Hobart AARNet networking conference in December 1991, that network traffic measured at the hubs has grown from 40 gigabytes in January 1991 to 120 gigabytes in late September. FTP is the largest user of bytes with 100 million packets per second. AARNet is now linked to Lae in Papua Hew Guinea, and will soon be connected to Part Moresby. There is also a Thai gateway. AARNet services thirty eight Australian higher education institutions and twenty four CSIRO divisions. AARNet reports to the AVCC Standing Committee on Information Resources which also has reporting to it library and computing services at a national level, In 1992 the US link has been upgraded and AARNet's service role will be enhanced. Interfaces to the public networks such as Austpac and Dialcom will be introduced necessitating only one terminal to access all information services. Up-to-date news on AARNet can be obtained from the AARNet office.[4]

For the individual users, of course, the technological underpinning of the network is irrelevant. What linked networks allow them to do, as it does elsewhere in the world, is to access information sources worldwide, in many instances without charge. Thus scholars at ANU can access the myriad of databases on the INTERNET, access document supply services like the Uncover service of CARL (Colorado Alliance of Research Libraries) - although at the time of writing CARL has been extremely tardy in setting up an automatic international standard fax facility to users overseas. Again, the medium is often there but people with their messaging are tardy!

Similarly, the recent upgrade of access to the British network JANET allows Australian access to many data resources such as the Oxford Text Archive. One of the major problems is educating the potential users of the services. Even people used to electronic mail are not aware, unless shown or made aware through easily accessible menus, the range of materials that are available. A recent study[5] at Murdoch University in Perth on AARNet highlighted the great need for the promotion of AARNet and training in a survey of its academic and general staff. While they found that 55% of academics had access to AARNet form their work place and 55% FROM home via a modem, only 20% were actually using AARNet, while 86% indicated they wanted to use AARNet. There is a need for marketing demonstrations, the provision of manuals, and electronic help desks on campuses.

The University of Newcastle Library has been extremely successful in putting together AARNet packages including floppy disks tailored to individual departments such as the English Department, in order to involve traditionally reticent humanistic departments. The University of Newcastle had found that only 15.6% of their academics surveyed in 1991 were using AARNet and only 17.5% of respondents were aware of the services provided through AARNet.[6] It is fascinating as to who sees their role in providing these AARNet services, certainly not always the Directors of Computing Services - it has been left in a number of instances to the university libraries in Australia to provide the background and training to such advances.

National and Regional Library Networks in Australia

The Australian Bibliographic Network (ABN) was established by the National Library of Australia in 1981 as a national shared cataloguing system. It now provides the basis for the National Bibliographic Database (NBD), which contains over fourteen million holdings records in January 1992. Essentially a finding tool as well as a cataloguing source, its technological infrastructure is currently under review to allow more flexible interaction, improved hardware and software, etc. It is basically a system relying on 1970's technologies with main memory and online storage. Hence software changes are difficult to implement and it is tied to the IBM main frame environment. Its current redevelopment programme undertaken with the National Library of New Zealand will see radical changes introduced in technological infrastructure. The NLA and RMIT's (Royal Melbourne Institute of Technology) INFORMIT hope to issue a union list of serials on CD-Rom in the near future based on the NBD.

Regional networks such as CAVAL (Cooperation Action by Victorian Academic Libraries in Victoria) and UNISON (University Libraries in New South Wales) provide the basis for multi-catalogue access by either combining databases in the former or linking them with a common software in the latter. An excellent overview of both network and individual library automated developments has recently been provided based on contributions prepared for the November 1991 Victoria Association for Library Automation Conference.[7]

Campus Wide Information Systems

These are less well developed than the norm in the United States but usually in advance of the United Kingdom. Much has depended on the efficiency of across campus contacts, for example, between Computing Service Directors, Head Librarians, and relevant Central Administration Personnel. Most will follow the example of the University of Melbourne and adopt WAIS (Wide Area Information Servers) systems as the structure for the delivery of their campus information.

National On-Line Initiatives

The CAUL (Committee of Australian University Librarians) group made a bid in 1991 to mount ISI (Institute of Scientific Information) databases on a host computer to service academics nationwide through AARNet. This was similar in principle to the JANET initiative funded by the Universities Funding Council in the United Kingdom. Due to some confusion within the bureaucracy of the Federal Department of Employment, Education, and Training (DEET) only one third of the required sum bid for was granted. At the time of writing it is hoped to purchase the equipment and mount a smaller sample of ISI or another database to prove the viability of the national network approach for individual academic terminal access.

DEET in 1991 made several major library related database grants, some of dubious validity but others such as the marketing of the Japanese Nikkei database within five universities being very important within Australia's business and economic infrastructure context. The Australian National University Library negotiated in 1991 with Reuters an promotional access package which has allowed significant free usage of the Reuter's database for one year which is another evidence of 'easing' the introduction of charged access mechanisms. Similarly this University launched its International Economic Data Bank database in 1992 under the title STARS (Statistical Retrieval System) with the World Bank, UN, OECD, and other relevant data being available on the local area network.

CD-Roms

CD-Rom networking on local networks is increasingly popular although many universities face the well known problem of cancellation of hard copies to cover costs of CD-Rom on line purchase. User habits and traditions are most important factors in the politics of change. At the Australian National University Library A$20,000 has been made available in 1992 to fund trial subscriptions to allow users to become familiar or be 'lured' into usage. In specialist areas such as Law CD-Roms additional funds have been made available to

purchase discs for network use. It is often easier for university administrators to see the value of library services in a network environment!

A national variant is to collectively buy CD-Roms. Thus CAUL negotiated in 1991 a discount deal with Chadwyck-Healey Ltd to purchase the English Poetry Full-Text Database for twelve university libraries. The use of non-copyright texts incidentally in this database raises interesting intellectual issues for the future as students browse allegedly textually flawed poems, while more scholarly analyses and textual versions sit on the traditional library shelves unused.

Electronic Short Loans/Scanning and Document Transmission

Short-loan (closed reserve) scanning developments have been hampered by lack of powerful equipment on campus or local bureaus able to cope with the needs of textual, mathematical, and cartographic material. The leading developments have been undertaken at RMIT with its Electronic Document Collection Program.[8] Copyright clearance has been sought by the Australian National University from academic authors on campus who hold their own copyright. Moving beyond this, trial users will probably see schemes similar to those being established by the State University of California at San Diego.[9] A number of universities have ordered the ARIEL RLG software and associated hardware and will introduce them in 1992 for local and international document sharing and supply.

Staff Attitudes and Access Philosophies

The actual switch in traditional large libraries form storage to access philosophies is difficult for staff to envisage, perhaps more so that the introduction of automated systems into libraries. Staff are used to working in traditional work flows even if these may not be effective either on a life cycle costing basis or on a efficiency basis. The types of workplace changes identified by Sue Martin in the working group on 'Strategic Visions for Librarianship' reflects the two types of librarians: those who are on the net and those who are not on the net, those who are interested in information access of this kind, those who are not.[10] The tensions will not diminish in the debate on priorities and service mechanisms in the 1990's.

Document Access and Supply

In the wider dimension libraries have to demonstrate that they can deliver documents efficiently and within reasonable costing frameworks to the user directly. In this area Australian libraries have been lagging behind. Australian users can go through to the CARL Uncover service and look at the contents pages of many Australian journals which are held in that group of libraries. Nowhere in Australian can one go to such a central source to look at the contents pages of Australian journals online in the same way. Access to the National Bibliographic Database is complicated for the user and has subscription and usage costs. Thus we have the paradox of Australian users being able to dial into MELVYL and the British Library, but not being able to dial into the National Library of Australia's own catalog not the National Bibliographic Database. The National Library of Australia is aware of these problems but as it was not on the AARNet until recently, nor had it a sophisticated internal microcomputer network, it has not had time to reflect upon the intellectual dimensions and problems that accrue from the global interconnectivity. When they reflect upon the potential for national linking in this area significant changes in 'outreach' beyond their present programmes might occur.

The major concentration of research material and budget now resides with the libraries of the universities. The State Libraries have, by and large, outside of their regional Australiana groupings, been forced, by budget considerations and other political pressures, to opt out of this particular sphere. The National Library's acquisitions budget of just over A$6 million, plus legal deposit, is only marginally more than the leading University Library, the University of Melbourne Library, so that their role as a national document supply centre is limited. Fifty percent of all Australian interlibrary loans are knock-for-knock, i.e., they are "free" so the

need to go to the National Library or to other suppliers is influenced by cost factors. Nonetheless it should be remembered in the electronic environment somebody somewhere has to pay for information- it is either subsidised or recouped.

Australian Higher Education Libraries

The Australian higher education scene is extremely complex at the moment.[11] The dissolution of the binary divide has meant that there is a sudden 'creation' of new universities. Many of the former colleges of advanced education had specific vocational educational roles but now aspire to research, without having the necessary infrastructure or increased funding to accommodate this. The end result is a pooling of poverty in which the large libraries are declining in real terms. The smaller libraries are aspiring to provide adequate material for undergraduate courses but are also entering into courses which are highly library intensive, such as Law, and for which there are not significant electronic alternatives to hard copy in Australia.

The establishing of a CAUL Working Party on Networked Information in 1991 with the following terms of reference is relevant here. It aims:

1. to provide a focus for the systematic development, and dissemination, of knowledge about the networking of information which can be used to support the activities of CAUL and its members;

2. to formulate strategies which enable CAUL to assume a leading role in the development of information technology networking policy relevant to the higher education sector;

3. to develop proposals which enhance access to information by the co-operative use of network facilities;

4. to represent the needs of university libraries to the publishing, computing, and telecommunications industries;

5. to promote effective use of networked information and facilities;

6. to coordinate joint projects.

There also needs to be, however, a heightened awareness to bring in the academic community in such deliberations. There is no equivalent in Australia to groups such as the Coalition of Networked Information. It is therefore reassuring that the Australian Academies (Humanities, Social Sciences, Science, and Technological Sciences) have taken the initiative in 1992 in planning a major conference on scholarly communication which will take place hopefully in early 1993, to bring together public service, educational, information specialists administrators, and users who will deliberate the changes in scholarly communication. This is scarcely new in the United States[12] but in an Australian context this could provide some useful local perspectives.

Thus we have Australia as a microcosm of the world scene; very much connected with the network developments; very much involved in accessing electronic information; very keen to have material delivered directly to users; very interested in involving themselves in developments in areas such as CJK script automation but only slowly realising its obligations re Australiana and the new modes of direct user interaction with document supply agencies. A great potential also exists in the supply of information from and to the Asis/Pacific region but this is also not yet realised. The new marketing play of the Australian International Development Program of library and information services to Asis could prove a useful catalyst in this area.

Distributed National Collection of Electronic Information Access

The concept in Australia at the present time of a Distributed National Collection[13], that is, the collected sharing of the distributed resources, will only become effective if the electronic networks are utilised in a truly co-operative but realistic sense. The vast continent of Australia is losing its internal as well as external 'tyrannies of distance'. We don't have the answers as to how this will all evolve but at least we are now able to play the game!

---

(1) Colin Steele. Satellite Television and Information Access in Australia" The Electronic Library, 8(1990), 181-184

(2) For example, Reva Basch, "Books Online", Online (1991), 13-23; Brett Butler, The Electronic Libary Program" Library HiTech, 34(1991), 21-30; Colin Steele, "Academic Libraries" in Librarianship and Information Work Worldwide. Edited by M. Line (London: Bowker-Saur, 1991) pp 15-46.

(3) Bruce Juddery. "Around the World by AARNet". Australia Campus Review Weekly, October 24-30, 1991, 11, 14.

(4) AARNet information can be obtained from Geoff Huston, AARNet Technical Manager, e-mail: aarnet@aarnet.edu.au

(5) Neil Greeve and De Stanton. "AARNet Survey 1991 - Murdoch University and Curtin University of Technology". Paper presented to the AARNet Networking Workship, Hobart, Tasmania, 1-3 December 1991.

(6) Jim Cleary. "The Role of the University of Newcastle Libraries in the Marketing and Instructional Support of AARNet". Paper presented to the AARNet Networking Workship, Hobart, Tasmania, 1-3 December 1991.

(7) M. Metcalfe ed. "Vala 1991: State of the States and Territories" Australia Library Journal, 40(1991), 270-288.

(8) Don Schauder *et al*. "Acquisition, Storage and Delivery of Information in Electronic Form" in Opac and Beyond. VALA 6th Biennial Conference. (Melbourne: VALA, 1991). Unnumbered report.

(9) Richard Goodman. Projects Currently Being Planned or Undertaken by the San Diego State University. (San Diego: State University Library, 1991.)

(10) Sue Martin. "Strategic Visions for Librarianship" Pacs-L message, 2 January, 1992.

(11) Colin Steele ed. Australian Higher Education Libraries. Adelaide: Auslib Press, 1992 (forthcoming)

(12) Richard M. Dougherty and Carol Hughes. Preferred Futures for Libraries. (Stanford: RLG, 1992.)

(13) John Thawley and Philip Kend ed. The Distributed National Collection. (Melbourne: ALIA, 1991.)

ELECTRONIC MENTORING OF LIS RESEARCH UTILIZING BITNET: AN ACRL PILOT PROJECT

Vicki L. Gregory
School of Library and Information Science
University of South Florida

On July 1, 1991 at the American Library Association Conference in Atlanta, the ACRL Research Committee launched a pilot project to mentor academic librarians in their conduct of research. The project was conceived while the committee was chaired by Charles Townley. Since the mentors and proteges are from all over the United States, the decision was made to mentor using the electronic conferencing facility of BITNET with the mainframe computer at New Mexico State University serving as the host LISTSERV machine. The actual use of the electronic conferencing facilities began on July 8 when most participants could be assumed to have returned to their offices.

Mentors and proteges are divided into six groups based on their subject areas of research: bibliographic control, collection management, expert systems, library effectiveness, scholarly communication, and understanding the user. (The Library Effectiveness Group actually functions as two subgroups to accommodate the number of proteges interested in participating.) Each group operates as an electronic conference with messages distributed by the LISTSERV computer to each participant of the particular list. Members of the groups also have access to everyone's electronic mail address by means of the directory provided in the project manual; therefore, participants can send private messages to a particular mentor or protege, but participants are encouraged to send all research-related communications to the group. Each electronic conference has a facilitator, from two to four mentors, and up to 20 proteges. Overall, the project has about 110 participants, counting mentors and proteges. It was decided not to moderate the conference in the customary way because of the inherently fragile nature of the mentoring process.

In this pilot project, beginning researchers have the opportunity to work with several experienced researchers from all part of this country. Electronic conferencing eliminates the problems of telephone tag and differences

in time zones; to carry on a coast-to coast mentoring relationship would be stressful, to say the least, using conventional means of communication.

Rationale for the Project. Many beginning researchers feel isolated from any sources of help. Their own library may have no one else interested in doing the same type of research as they are, or even have no one interested in doing research at all. This pilot project is intended to put beginning researchers in contact with mentors and peers who can be of assistance to them both now and in the future.

Although many of the beginning researchers will have taken advantage of some of the continuing education courses in research that ACRL and other library professional organizations have offered as pre-conferences and workshops across the country, and there is doubtless a great deal to be learned from such "one-shot" meetings, the members of the ACRL Research Committee trust that a longer relationship with mentors will prove to be an excellent complement to these workshops and provide for a more individualized approach to problem-solving and other issues in library research. An experimental element in this project is the concept of group mentoring. Almost by definition, mentoring is considered to be a one-on-one experience. Thus, one very important objective of this project, and a key to its success, is to foster the development of the kind of trust that mentoring requires in a group situation maintained, for the most part, by electronic means.

Early Analysis of the Pilot Project. To date the amount of traffic on the electronic conferences has been a disappointment to the planners of the project; however, some research-related mentoring has occurred in all of the electronic conferences. The participants in the Understanding the User Conference have initiated a group research project in the area of interlibrary loan usage.

Initially, discussion of changes in communication styles/media of the participants had been targeted as a potential topic for this paper. However, not

enough sustained use has occurred for this type of discussion to be possible except for those participants who have been active members of other discussion groups as well as one of the ACRL electronic conferences. The major issue to be addressed seems to be why more participants have not been active users, as opposed to simply readers, of the electronic discussion groups.

One obvious explanation for the slow start of the project is that all of the electronic conferences experienced system problems beginning in mid-August 1991 (roughly six weeks after the project was launched) when the "list owner," Thomas A. Peters, changed institutional affiliation. For about 6 to 8 weeks after his move, many or all of the messages sent to the listserver were not redistributed to the participants. Since Peters was still recognized by the host computer and could add, delete, and change addresses of participants, the system problem was not recognized for some time. After some complaints about "non-responses" to messages, the existence of the problem was suspected, and experimental messages were sent out with individuals receiving the message requested to send an e-mail message acknowledging their receipt of the message. When it was determined that there was definitely a system problem, Peters quickly re-signed everyone onto their respective lists, an action which seems to have cured the system problem.

Other than system problems, logical deduction suggests, as has been documented by research concerning other electronic conferencing projects, that there needs to be a critical mass of people in each of the groups. Markus states a critical mass of people communicating using the same medium tends to increase the per person frequency of communications.[1] The electronic discussion groups in the ACRL pilot project were set rather small, attempting not to have more than 20 proteges for every two to three mentors. The reason for setting such a small limit was that it was felt that the burden might otherwise be too great for the mentors. And, in fact, when the ACRL Research Committee members were recruiting mentors, the number of proteges in a group was a major concern for the potential mentors, some of whom thought that 20 proteges would be far too many. What we may have all failed to take into consideration is an old rule-of-thumb, which in

this case might be reworded to state: twenty percent of the members of the discussion lists would do eighty percent of the actual discussing.[2] With such a small number of participants in each group, the number of people actively sending messages to the list does not reach the critical mass needed to motivate active participation of more members of the group. Sproull and Kiesler state that if few people actually use the new form of communication, potential participants are driven away because they do not know if others are really getting their messages and have less incentive to send more messages.[3]

Another possible cause of problems with these electronic discussion groups may be inherent in the type of discussion required for mentoring and furthering research efforts. Prior research on electronic mail communication has found that electronic mail has been judged by participants to be especially useful for non-task-oriented objectives. In one such study, respondents to a survey thought that electronic mail was a very useful method of exchanging information and asking questions, but when more complex communication was required such as debate, discussion, and resolution of problems the perceived usefulness of the system declined.[4] The type of discussion planned for the LIS electronic discussion groups falls into the area of complex communication; and thus it may or may not ultimately prove to be appropriate for this medium of communication.

Although not represented by the electronic survey discussed below, another related problem, known through discussions at the meeting of the groups in Atlanta, is that a significant number of participants are first-time users of BITNET or Internet. Depending upon their local computer center, they may be experiencing what one writer quoted in Forbes magazine called Internet's "savage user interface."[5]

Project Meeting in San Antonio. At the 1992 American Library Association Midwinter Meeting (roughly 6 months into the project), 25 participants in the electronic conferences attended a session where they were given an opportunity to meet in three groups, based upon their area of research interest,

to discuss the project and their research. After these discussions, a recorder from each group reported to everyone concerning the suggestions and concerns of their group. Several concerns were expressed which would have an impact on their electronic communication patterns.

The major problem expressed was, as I would describe it, a general shyness. Several people expressed the idea that they did not feel that their research was significant enough to communicate with others; however, after face-to-face discussions and encouragement, these participants said that they now felt that they would be more comfortable in the future to pose questions and make comments via the electronic conference. Also some attendees thought that more structure and direction from the mentors would encourage them to overcome their hesitancy to communicate via the electronic conference.

One group mentioned the technical problems experienced early in the project had discouraged participation for the number of people who had sent messages and gotten no response. Participants were assured that the listserver was now functioning properly and that the conference messages would be distributed to the members of each list.

Another concern expressed by several participants a fear of someone in that anonymous-feeling electronic environment stealing their research idea. Members of the ACRL Research Committee tried to persuade the proteges that since each group had roughly 20 members that there was protection in that number of people reading the messages. However, this fear is not just a product of the electronic environment, but a common, if mostly unjustified, worry of the beginning researcher. Additional opportunities to meet face-to-face with others in their electronic conference will hopefully help to overcome some of these kinds of concerns.

At the end of the session, the overall feeling of the attendees seemed to be positive toward the continuation of the project and optimistic concerning future activity on the lists.

Electronic Mail Survey of Participants. In an attempt to get some information about the experience and use of electronic mail/conferencing by the participants, a survey was sent out to all participants via the listserver, soon after the system problems were resolved in October. Out of a possible 108 respondents, 36 or about 33 % replied to the survey, either by electronic mail or, after printing off the questions, by regular mail.

Of those responding to the survey, all but one had had a computer account to access either/or BITNET/Internet before joining the project. However, I strongly suspect that many of the nonrespondents obtained an account because of the project, and lack of experience with the system may have inhibited or even prevented these people from responding to an electronic mail survey. For the respondents, the mean number of years of experience they have had with electronic mail was 2 years with the responses ranging from a high of 6 years to a low of 5 months. Thirty-two out of a possible 36 belonged to other electronic discussion groups in addition to the ACRL mentoring group, with the average respondent belonging to 3 additional electronic conferences. The discussion group that was most mentioned was PACS-L to which 15 of the participants belonged, followed by AUTOCAT, LIBREF-L, and LIBADMIN with 7 each. Overall, respondents belonged to 42 different discussion groups. Approximately 69% of the respondents to the questionnaire had a terminal on their disk at work from which they could access BITNET or Internet. Only three participants did not have access to a terminal in their immediate work area although another did report that because of the number of people competing for use of the terminals that it was difficult to get time at the computer in the immediate work area. Thirty-nine percent of those responding were able to access the network from home and 61% could not.

Most of the respondents to the survey (78%) felt that the use of BITNET and Internet for electronic mail was easy once they learned how to use it. A common complaint was a lack of documentation on basic commands and procedures which contributed to the large amounts of time required to become proficient in its use. Many indicated that colleagues in the library had been particularly

helpful in teaching them about BITNET/Internet, with the library's local systems or automation librarian conducting much of the training that was necessary to learn the system. Forty-seven percent of those responding indicated that they felt that the local interface for BITNET/Internet was user-friendly, about 36% felt it was not user friendly with the rest in a "somewhat" user friendly category.

When asked as to the major problem experienced with BIT- NET/Internet, several areas were identified by about 57% of the respondents which could, it appears, be obviated or at least ameliorated through better local system documentation. Several respondents experienced technical problems using their microcomputer from home in respect of not knowing how to correct errors, or even to clear the screen, problems which could be solved through making more local documentation available to users. Five respondents indicated problems in understanding electronic mail addresses and problems in knowing how to send electronic mail messages to users of networks other than their own. Twenty-eight percent of the respondents indicated problems in knowing how to edit and send files again where proper procedures were inadequately, or not at all, documented. Five Internet users indicated that they had experienced problems using the FTP function to access files in other (nonlocal) computers.

Another problem area identified had to do with local system capacity. Six respondents indicated that except in the early morning or evening it was difficult to get a port into their local computer. Some others experienced response time problems, making reading and responding to electronic mail a very time-consuming affair. One person indicated that logon procedures took, he thought, an "incredibly long" time because of the number of screens involved and the time spent waiting for a response back from the computer.

Ninety-one percent of the respondents felt that having access to electronic mail had changed their communication patterns. Many indicated that their preferred medium of communication had become electronic mail. Seventy-eight percent of the respondents said that they were better able to communicate with colleagues at other institutions than before they had access to

electronic mail. Many mentioned that electronic mail had eliminated the problems of "telephone tag" and allowed them both to ask and receive answers much more rapidly. Widening their field of professional acquaintances and even making friends over the electronic discussion groups was mentioned by several participants. Electronic mail was seen as a replacement for writing memoranda and letters and a replacement for many of the telephone calls that they would have made in the past. One participant mentioned that many questions, which would otherwise have had to wait for professional meetings and perhaps not have been answered there, had been answered for her over the electronic discussion groups,

Many participants indicated that they used electronic mail to communicate with colleagues in their own library and on the same campus. One branch librarian mentioned that access to electronic mail had made her feel more a part of what was going on in the main library than she ever had before. One respondent indicated that electronic mail was very helpful in keeping his supervisor informed of various projects at different stages. This person reported that he reports "minor" problems and concerns over electronic mail in order that departmental meetings can be devoted to "larger" issues. While some research in other fields has indicated that use of electronic mail stimulates the use of all media of communication, including the telephone,[6] this phenomenon was not mentioned by any of the respondents to the questionnaire.

Future Plans. Based upon information obtained in the survey and from the group discussions in San Antonio, some restructuring of the conferences is being examined. One suggestions is to collapse the library effectiveness groups into one electronic conference with roughly 40 participants. Consideration is also being given to establishing an overall discussion group which would include all participants that could be utilized for general (non-subject specific) research questions and comments.

The ACRL Research Committee is planning a program meeting for the 1992 American Library Association Annual Conference in San Francisco which will

consist of an evaluation of the pilot project from the standpoints of at least one mentor, one protege, and one committee facilitator. At the end of the presentations, some time will be devoted again to meetings of the electronic-conference groups.

If the project is deemed to have been at least reasonable successful by the time of the San Francisco meeting, the project will be continued, probably in an expanded mode. International participation is being considered, as BITNET through various gateways offers access to institutions around the world. New groups or areas might be established or some groups might be consolidated depending upon the experiences and input of the project participants.

Assuming continuation of the project, the intent of the ACRL Research Committee is to open up the electronic discussion groups to more participants, expecting that over time, certain combinations of proteges and mentors will break away from the big group as contacts are made electronically with individuals that share similar research interests.

In closing, it should be remembered that some mentoring has taken place in all of the electronic conferences, and the ACRL Research Committee is optimistic that as time goes on more message activity will be taking place. This type of mentoring environment may not be appropriate for every beginning researcher, but, for those willing to give it a try, great opportunities exist for fruitful guidance and encouragement from experienced researchers and other proteges from all over the country.

NOTES

1. M. L. Markus, "Toward a 'critical mass' theory of interactive media: Universal access, interdependence and diffusion, Communication Research 14 (1987): 491-511.

2. This general maxim has been shown to be the case in other electronic discussion groups, see, for example, Dave Cook and Michael Ridley, "Computer-mediated Communication Systems: Will They Catch on?" Canadian Library Journal 47 (December 1990): 413-417.

3. Lee Sproull and Sara Kiesler, Connections: New Ways of Working in the Networked Organization (Cambridge, Mass.: MIT Press, 1991), p. 166-167.

4. Dave Cook and Michael Ridley, "Communication-mediated Communication Systems," Canadian Library Journal 47 (December 1990): p. 414.

5. David Churbuck, "Civilizing Internet," Forbes, July 8, 1991, p. 90.

6. Tony Kaye,"Introducing Computer-Mediated Communication into a Distance Education System," Canadian Journal of Education Research 16 (Spring 1987), p. 157.

DESIGN FOR AN ADAPTIVE LIBRARY CATALOG

by

Michael K. Buckland, Barbara A. Norgard, Christian Plaunt.
School of Library and Information Studies,
University of California, Berkeley, CA 94720

Abstract: A progress report on the design and demonstration of a prototype adaptive online catalog, OASIS. Online catalog searches commonly retrieve too few or too many items. This prototype, implemented as a transparent, workstation-based front-end to the University of California's MELVYL tm online catalog, adapts to excessive or insufficient retrieval by strategically limiting, sorting, or expanding users searches, based on preferences defined by the user.

BACKGROUND

Libraries are installing online catalogs, which are becoming larger, more powerful, and require increasing expertise for effective use. Most library users have little familiarity with the Library of Congress Subject Headings or with the MARC format (Butkovich 1989, Peters & Kurth 1991). Library users have no choice but to use the online catalogs, where installed, yet are not expert searchers and use few of the available commands (Bellardo 1985, Shenouda 1990).

Each user's search is unique but it has seemed helpful to categorize the quantitative outcome of search results somewhat arbitrarily as follows:

Category	Number retrieved	User's need
Zero	None	Find something
Too few	1-4	Find more
Acceptable	6-15	Satisfied
Too many	16-500	Find fewer
Far too many	>500	Find fewer

This research is concerned with the twin problems of retrieving too few or too many records. The solution being adopted is the use of strategic search commands within the broader concept of an adaptive retrieval system.

"ADAPTIVE" RETRIEVAL SYSTEMS

Our working assumption is that retrieval systems should be designed to be adaptive. By this we mean that, in principle, no matter what the user's query may be, no matter what data the database may contain, the system should be designed to supply the the preferred number of the desired kind of records. Bibliographic retrieval systems have traditionally been designed to retrieve everything matching some specification, but it is rare that anyone does want everything and, further, the search specification used is normally a very incomplete representation of all the searcher's specifications. For example, users ordinarily have a preference for relatively up-to-date material in languages that they can understand, but this is rarely specified explicitly. An adaptive online catalog would be one designed to help the searcher to adapt the search in relation to their preferences as their search evolves.

STRATEGIC SEARCH COMMANDS

Expert, effective searching of online bibliographic systems is done by implementing a search strategy composed of a series of tactical moves. In practice, however, not all searchers are expert. Weak expertise is

associated with a lack of knowledge of search commands, of search strategies, and of how the material is arranged in the database. Weak expertise is a significant problem in the case of online library catalogs, which are used by untrained searchers. As the functionality of online catalogs increases, so their complexity increases, and, so too, the amount of expertise needed for the task of using them.

Libraries are replacing card catalogs with online catalogs. Library users have no option but to use the online catalog. Examination of online catalog usage indicates that very few of the available commands are frequently used. In particular, as the size of the files grow with the retrospective conversion of older records the frequency with which excessive numbers of records are retrieved increases. Expert searchers know of search tactics that can be employed to reduce retrieved sets. The great majority of relatively inexpert users typically scroll through page after page of displayed records, settle for the first few found, or start over with some new search command (Walker 1990).

Increasing complexity in a self-service system leads, unless remedied, to an increasing discrepancy between the prevailing level of expertise and the expertise needed for the task. The remedy being explored in this research and demonstration project is to enable the system to supply some of the expertise in tactical moves that an expert human intermediary would supply. If the searcher can provide direction, the system should assist in moving in that direction. Automatic transmission in automobiles provides a suitable analog. The driver decides when to move, in which direction, and at what speed: The automatic transmission shifts gears as appropriate. It is important to note that the technical complexity of shifting gear has not been reduced by automatic transmission. Rather, some of the technical complexity has been delegated to the transmission system. It is the complexity of the task facing the driver that has been reduced (Buckland and Florian 1991).

In the spirit of search strategy analysis by Bates (1979, 1990), Fidel (1985), and others, we use the term "strategic search command" to denote a search command that instructs the system to implement a series of tactical moves in some direction. Given the propensity of library users to limit themselves to only a few commands, it is difficult to see how else increasing complexity can be handled except by providing more versatile commands. We define a strategic search command as a command that implements a series of tactical moves which could be taken separately. As with automobile automatic transmission it is a matter of enabling the user to delegate some of the complexity to the system and, as with automatic transmission, it is necessary that the user remain in control of the pace and direction. What works for the non-expert is also likely to be a convenient amenity for the expert.

Strategic search commands can be identified for problematic search results as noted above. This project addresses three:

Category	Number retrieved	Optional Strategic command
Too few	1-4	MORE
Too many	16-500	FEWER
Far too many	>500	FEWER

FEEDBACK TO THE SEARCHER

The use of automatic transmission does nothing to reduce the need for the automobile driver to have an excellent understanding of local geography and of road and traffic conditions. Good automatic transmission enables the driver to concentrate more on navigation and less on the mechanics of driving. Similarly, delegating search tactics to an online catalog should allow the searcher to concentrate more on understanding the "terrain" and more on navigation.

Effective navigation depends on adequate, reliable, intelligible information concerning the options available. Delegation of tactical moves to the system does nothing to reduce the need for informative feedback on the search situation. Consequently design objectives include: (i) increasing the searcher's understanding of the search status and retrieval situation; (ii) presenting the search options; and (iii) indicating what the system suggests doing next. The intention is not only to empower the searchers's ability to do things but also to

enhance the searcher's ability to direct searches knowledgeably.

PROTOTYPING USING A FRONT-END

In the development of complex systems, it can be very useful to construct prototypes in order to demonstrate the effects of interesting, alternative approaches. The experience derived from experimenting with a prototype can then provide an informed basis for developing a robust "production" version for routine use. Prototyping is a form of experimentation that does not replace, but complements, both analytical studies and the design of complete production systems.

The development of online catalogs with meaningfully large databases is a very expensive undertaking even for a prototype. A very economical approach is to use a second computer to add functionality to an existing operational system. In the work reported here a Unix workstation (DECStation 5000/200) is used as front-end to the MELVYL catalog. MELVYL is a "second generation" online union catalog which provides access to some 8 million catalog records representing the holdings of the hundred libraries of the nine campuses of the University of California. An experimental front-end, Otlet's Adaptive Searching Information Service (OASIS) operates routinely as if a dumb terminal providing transparent access to MELVYL. However, by recognizing, intercepting, and performing local operations on selected commands and search results, the OASIS prototype provides additional functionality as if provided by enhancements to the MELVYL system itself.

EXCESSIVE RETRIEVAL

We can illustrate the present problem of excessive retrieval with a simple example. In MELVYL, a search on the subject of the Algerian Revolution might well be expressed as a subject keyword search FIND SUBJECT ALGERIAN REVOLUTION. This search yields 133 records, more than is likely to be wanted and more than is convenient to browse. Since MELVYL follows the tradition of card catalogs in presenting records in alphabetical order of main entry, the first few found are not likely to be any more interesting than any others.

An expert searcher could reduce the retrieved set by trying various search modifiers. For example, in this case:

(1) AND AT BERKELEY reduces the set to 64, all more conveniently located for a Berkeley-based searcher but still a lot to browse through.
(2) AND LANGUAGE ENGLISH reduces the set abruptly to only 11.
(3) AND LANGUAGE ENGLISH OR FRENCH makes little difference, reducing the 133 only to 119.
(4) AND DATE SINCE 1983 would exclude the older half, but this particular date limit is not currently allowed on MELVYL and, even if it were, it is a tedious chore to determine that it is the year 1983 that needs to be chosen in order to halve the retrieved set.

As practice, a combination of search modifiers is likely to be preferable to any single search modification. For examine, most Berkeley-based searchers might find attractive the combined effect of (1), (3), and (4) which yields a convenient set of 14 fairly recent records of materials in English or French. all held on the Berkeley campus.

Note that the effects of any given search modification will vary for every single search, are difficult to predict, and are more or less tedious to ascertain, even for an expert searcher. The effects of any given combination of search modifications are all the more difficult to predict or to ascertain, However, since these are systematic variations on stored data, the system can be programmed to analyze and to report the effects of search modifications.

THE "FEWER" COMMAND

Two different strategic commands have been developed for reducing large sets: One, designed for minimal augmentation of MELVYL's capabilities, is called "FEWER"; the other, "FILTER", makes substantial

use of the front-end.

The FEWER command uses the search modifications supported by the MELVYL system. In the OASIS front-end a small list of MELVYL-intelligible search modifiers is stored. The current default list is:

AND AT BERKELEY
AND LANGUAGE ENGLISH
AND DATE RECENT [last 10 years]
AND DATE CURRENT [last 3 years]
AND FORM BOOK

The new command "FEWER", unintelligible to MELVYL, is offered as if a MELVYL command but is intercepted in the front-end which substitutes and forwards in its place the first (or next) in the list of MELVYL-intelligible search modifiers. A user encountering too large a set can enter the command FEWER repeatedly for progressive reduction of the retrieved set. OASIS intercepts the FEWER command, substitutes the first (or next) search modification, with cumulative effect, forwarding to the user the effect of each modification. For example, a subject keyword search on "Napoleon" yields 4,580 records in MELVYL and one on "Libraries" yields 16,613 records. The effects of the repeating the FEWER command are:

Commands	System implements	Resulting set size	
		Napoleon	Libraries
FEWER	AND AT BERKELEY	2,259	10,346
FEWER	AND LANGUAGE ENGLISH	853	8,543
FEWER	AND DATE RECENT [last 10 years]	73	2,837
FEWER	AND DATE CURRENT [last 3 years]	28	830
FEWER	AND FORM BOOK	26	819

The user can change the default list to reflect personal preferences. The FEWER command is proving to be effective with very large retrieved sets.

THE "FILTER" COMMAND

The FILTER command draws more heavily on the front-end to augment MELVYL. Faced with a set that is too large but seems worthy of analysis, the user can issue the command FILTER as if it were a MELVYL command. The OASIS front-end intercepts the command, substitutes commands which make MELVYL transmit selected MARC fields for all of the records as if for continuous display at a dumb terminal. The OASIS front-end, however, intercepts this display data and, instead, stores it in memory to be analyzed in terms of the user's personal preferences.

The first experimental version assumes that, faced by an excessive retrieved set, the user will tend to have preferences with respect to:

DATE:	Newer preferred to older;
LANGUAGE:	English preferred to other; and
LOCATION:	Local (i.e. Berkeley) to distant holdings.

The front-end analyzes the retrieved set in terms of these three preferences, creates a three-dimensional array, and displays a summary to the user. For example, a MELVYL subject keyword search on "Dresden" yields 440 records. The FILTER command generates the following analysis:

Location:	Berkeley campus		Other campuses only		
Language:	English	Other	English	Other	Total
1990-91	0	5	2	3	10
1980-89	9	37	6	48	100
1970-79	3	25	12	54	94
1960-69	5	26	12	42	85
1950-59	0	15	0	14	29
1900-49	6	28	8	39	81
1800-99	1	9	4	18	32
1700-99	0	4	0	3	7
1600-99	0	0	0	1	1
Pre-1600	0	0	0	1	1
Total	24	149	44	223	440

In this initial version the front-end ranks the subsets of records to reflect a default preference for recent, English-language, Berkeley holdings. However, the user can override the default to select any subset for a display of the records. In the FILTER default display the records are displayed, subset by subset, mimicking MELVYL's short display format. The FILTER command is being developed to provide more flexibility.

TOO FEW

Implementation of a "MORE" command is in progress and we are experimenting with alternative approaches. The current design can be seen as a special case of the more general case of a directed FIND RELATED MATERIAL command. The searcher can ask for a display of the subject headings to be found in a search. In one version the Library of Congress Subject headings are excerpted and those most frequently are displayed, ranked by frequency of posting within the retrieved set. For example when applied to the 55 records retrieved by a subject keyword search on "BERKELEY AND RENT" the following display is provided:

1.	Rent control -- California -- Berkeley	75
2.	Berkeley Calif Rent Stabilization Board	16
3.	Elmwood District Berkeley, Calif	6
4.	Housing surveys -- California -- Berkeley	6
5.	Berkeley Calif -- Politics and government	4
6.	Rent control -- California	4
7.	Rent control -- California -- Berkeley	4
8.	Berkeley Calif -- Dwellings	3
9.	Berkeley Calif -- Statistics	3
10.	Housing -- California -- Berkeley -- Statistics	3
11.	Local elections -- California -- Berkeley	3
12.	Rent control -- California -- Berkeley -- Statistics	3
13.	Rent control -- California -- San Francisco	3
14.	Berkeley Calif Rent Stabilization Board -- Elections	2
15.	Elections -- California -- Berkeley	2

This display, which mimics a MELVYL display for browsing subject headings, provides an informative

basis for the searcher to extend the search selectively among related materials. It would be desirable to display in addition, the frequency of posting in the host database so that the searcher would have an estimate of the number of records that would be retrieved by using each subject heading.

Another option fragments the LCSH in the retrieved set into their constituent terms (words or phrases) and displays a similar ranked list. A more automatic approach had originally been intended but we have preferred to place more information and more "steering" in the hands of the searcher. A practical problem is that the most frequently occurring subject terms are, because they occur frequently, the ones most likely to generate excessive retrieved sets if used.

REFLECTIONS

Several comments can be made concerning the work being reported:

1. As with automatic transmission, automatic cameras, and other applications of artificial intelligence, the underlying strategy is to move complexity into the system. The complexity is not reduced, but some of it is delegated.

2. The FILTER command is of interest because it represents a two-stage approach to retrieval. The first stage is the initial search on a data base of 8 million records; the second on a subset of, say, 80 or 800. The second stage allows new options for retrieval techniques because computer-intensive techniques impractical for a very large set may be feasible for a small set. Further, the small set selected in the first stage for processing in the second stage will generally contain a much higher proportion of relevant records ("generality") than the database as whole. This increased proportion of relevant records changes the retrieval problems and options. A front-end can be programmed to support a variety of retrieval options not supported by the main system.

3. Subset ranking provides an alternative to the two orthodox methods of arranging retrieved sets: Ordering by alphabetical order of main entry; and mathematically-derived strict document ranking.

4. An emphasis has been placed on non-topical attributes of records, notably date, language, and location. It is believed that preferences for non-topical attributes are of more interest to searchers than has been generally recognized, especially when too many items are initially retrieved.

5. Techniques based on downloading records to the front-end could operate on any data in (or implicit in) the MARC record, an option not generally available in online catalogs. Downloaded records could be searched, for example, for place of publication and for specific phrases within titles.

6. Prototyping in this manner offers two possible by-products: A convenient opportunity to try out possible modifications to the main retrieval system at little cost; and a means of supplementing the main systems by providing a means for handling exceptional searches.

7. A two-stage, two-computer approach to retrieval was adopted for this work as the only economical approach to examining possible enhancements empirically. An advantage is that with the development of telecommunications and network protocols, such as Z39:50, a front-end can, in principle, be used with any online catalog that supports remote access.

Acknowledgement. The work reported in this paper summarizes a research and demonstration project supported by the US Department of Education under the Higher Education Act, Title IID, (#R197D00017), and by the School of Library and Information Studies of the University of California at Berkeley.

REFERENCES

Bates, M.J. (1979). "Information search tactics." Journal of the American Society for Information Science 30(4):205-214.

Bates, M.J. (1990). "Where should the person stop and the information search interface start?" Information Processing and Management 26(5):575-591.

Bellardo, T. (1985). "What do we really know about online searchers?" Online Review 9:223-239.

Buckland, M.K. & D. Florian. (1991). "Expertise, task complexity, and the role of intelligent information systems." Journal of the American Society for Information Science 42:635-643.

Butkovich, N.J and others. (1989). "An expert system at the reference desk: Impressions from users." Reference Librarian 23:61-74.

Fidel, R. (1985). "Moves in online searching." Online Review 9(1):61-74.

Peters, T.A. & M. Kurth. (1991). "Controlled and uncontrolled vocabulary subject searching in an academic library online catalog." Information Technology and Libraries 27:201-211.

Shenouda, W. (1990). "Online bibliographic searching: How end-users modify their search strategies." Proceedings of the 53rd Annual Meeting of the American Society for Information Science 27:117-118.

Walker, S. (1990). "Interactional aspect of a reference retrieval system using semi-automatic query expansion." In: Informatics 10: Prospects for Intelligent Retrieval. London: Aslib, 1990, 119-136.

Assessing Information on the Internet

Toward Providing Library Services for Computer-Mediated Communication

Martin Dillon, Erik Jul, Mark J. Burge, Carol A. Hickey
Office of Research
OCLC Online Computer Library Center

Abstract

This paper describes work in progress, funded by the U.S. Department of Education, Library Programs, to assess the nature of textual information available via the Internet, a global network of computer networks. An overview of the project's objectives and methods, preliminary findings, and a description of remaining work are presented.

Locating, accessing, and using information resources on the Internet, a global computer network of networks, can be difficult, time-consuming, and sometimes impossible. In this new and rapidly expanding electronic environment, network users have unprecedented access to information and computing resources. However, the development and implementation of systematic methods of describing and providing access to information lags behind deployment of the Internet itself, and the ability for network users to share information surpasses by far the ability to discover information on the Internet. Traditional library services such as cataloging have yet to find widespread application in this environment.

The OCLC Internet Resources project, funded by the U.S. Department of Education, Library Programs, investigates the nature of electronic textual information accessible via the Internet. This empirical study explores the practical and theoretical problems associated with providing traditional library services for electronic text in a wide-area network environment.

This paper describes the OCLC Internet Resources project and reports preliminary findings of the work in progress.

Objective

The primary objective of this project is to provide an empirical analysis of textual information on the Internet that will inform the efforts of those interested in cataloging or otherwise describing and providing access to electronic resources in a wide-area network environment.

Methods

Project methods include (1) locating, collecting, and analyzing a sample of textual information on the Internet, (2) developing and testing a taxonomy of electronic information based on the sample, (3) identifying and analyzing problems associated with cataloging, indexing, and providing appropriate levels of access to this information.

Document Collection

The early focus of the project is to collect sample text documents from Internet sources. Project staff use an array of resources to discover the whereabouts of electronic text, including printed books, journal articles, and newsletters; online electronic publications and lists; information discovery tools such as WAIS (Wide Area Information Server) by Thinking Machines Corporation, Gopher by the University of Minnesota, and Archie by McGill University; hypertext programs; electronic conferences; e-mail; and online browsing.

Information Map

The characteristics, location, and methods of obtaining electronic information on the Internet derive, at least in part, from the nature of the Internet itself. As of January 1992, the Internet comprised 3,591 different networks supporting at least 727,000 host computers communicating via the common TCP/IP protocol (Transmission Control Protocol/Internet Protocol).

The TCP/IP protocol suite provides three primary application services: Telnet, File Transfer Protocol (FTP), and electronic mail (Simple Mail Transfer Protocol, SMTP). Each protocol provides a distinct network function:

- Telnet allows network users to log on and communicate with a remote host computer as if there were a direct connection between them.
- FTP allows users to transfer files between remote and local computers. Many host computers maintain an accessible storage area for publicly accessible files.
- Electronic mail allows users to send and receive mail messages between host computers.

All three protocols enable the exchange of textual information; thus, systems and sites providing or facilitating access to text files using these protocols are of interest to this study.

Another major source of electronic information is BITNET, a "store-and-forward" network of IBM computers linking more than 3,000 host computers. BITNET protocol provides the following major application services that are relevant to this study:

- LISTSERV software manages the exchange of information using mailing lists. Many online conferences and discussion lists are based on this software, and many listserv sites archive data.
- Electronic mail enables users to exchange information directly with other users as well as to submit commands to remote network hosts.

There is a gateway between BITNET and Internet, and to the user, the network's boundaries are becoming increasingly less apparent. This project focuses on textual information available via the Internet regardless of the network of origin or the protocol used to disseminate or access the information.

The USENET network, an informal cooperative network which is heavily used to support information exchange within and among newsgroups, generates almost 31 Mb of information monthly. It is not reasonable to consider cataloging individual news items, and for this reason the text generated by USENET activity is generally beyond the scope of this project. However, USENET news does enter the Internet world through various feeds and archives. The problem posed by USENET relates more to cataloging archive sites and newsgroup sources than to individual texts.

A high-level overview of Internet resources, many of which are sources for electronic text, appears in table 1.

Table 1 Overview of Internet Resources

Type of Resource	Number	Internet	BITNET	USENET
FTP Sites	968	X		
FTP Files	2.1 million	X		
Listserv Sites	111		X	
Conferences	1,200	X	X	
E-Journals	26	X	X	
E-Newsletters	72	X	X	
E- Digests	16	X	X	
Other Servers (e.g., file, whois)	216		X	
Telnet Sites		X		
Library Catalogs	337	X		
Other	156	X		
Computer Centers	18	X		
USENET Newsgroups	1,157			X

Cataloging Emphasis

Using the primary Internet and BITNET protocols, project staff have collected approximately 1,200 text documents from various Internet sources including FTP sites, LISTSERV hosts, and interactive mail applications. For preliminary administrative purposes, the files have been separated into some 56 categories ranging from books and electronic journals to informal personal communications (table 2). (Although e-mail is beyond the scope of this project, some text files residing in publicly accessible directories are e-mail messages that have been saved. Determining this before retrieving the file is often impossible.)

Table 2 Profile of Document Sample

Abstracts	Directories	Lists	Quotations
Announcements	Documentation	Lyrics	Readme Files
Articles	Drafts	Manuals	Recommendations
Bibliographies	E-Mail	Minutes	Reports/Papers
Bills	Editorials	Monthly Reports	RFCs
Biographies	Encyclopedias	Newsletters	Standards
Books	Essays	Notes	Statements
Briefs	Fact Sheets	Poetry	Summaries
Brochures	Glossaries	Policies	Surveys
Bulletins	Guides	Press Releases	Theses
Charters	Hearings	Profiles	Testimonies
Conferences	Humor	Proposals	Tutorials
Dictionaries	Indexes	Public Laws	Weather
Digests	Journals	Publicity	Workshops

The initial document collection was created often as the result of directed searching, i.e., one document or information source would point to another. This introduces a bias into the collection, but does not prohibit preliminary analyses. To reduce the bias, a second collection will be gathered using automated methods developed by project staff.

Preliminary Analysis

One hundred documents from the collection have been selected for preliminary manual analysis. Information gained during this phase will assist development of software to perform automated document analyses.

Project staff examined each document to determine its characteristics and create a simple cataloging record. Not surprisingly, the completeness of information useful for cataloging the documents ranged greatly. Some electronic journals, for example, provided considerable descriptive data, including ISSNs (International Standard Serials Number), whereas other document types had none.

Of the one hundred documents, 96 provided some sort of information at the head of the file, before the text proper; 30 included additional information at the end of the file, following the text proper. Ninety documents had an identifiable title, but only 73 had an identifiable author, Fewer yet, only 64, had any kind of date within the text of the document.

Cataloging Initiative

Project staff are presently expanding the cataloging portion of this project. A natural next step is to apply the existing MARC (MAchine-Readable Cataloging) format for computer files to the documents in the collection. This exercise will reveal how effectively this format handles a broad range of electronic textual information, and will simultaneously reveal the degree to which these electronic text documents provide sufficient data for systematic cataloging.

The project staff, in coordination with professional catalogers and standards groups such as the Library of Congress and MARBI, will extend the cataloging initiative to discover, through repeated application of the format over a wide range of documents, the degree to which cataloging requirements are satisfied.

Overview of FTP Sites

FTP sites are a highly volatile source of electronic files of all sorts: text, data, and software. Not unlike much of the Internet, FTP sites represent a great terra incognita, and developing even a rough map of the files available through FTP sites was important if project staff were to gaining a better understanding of the textual information available via this Internet resource.

Approaching this task was greatly facilitated thanks to the efforts of McGill University and their Internet scavenger Archie, a software "knowbot" of sorts that visits Internet FTP sites periodically and retrieves a complete listing of files from each host. Each FTP site is visited approximately once a month, and data in the Archie database are therefore subject to this lag time. The file containing the listings of FTP sites is available from McGill University (host: archie.mcgill.ca.; directory: /archie; file: listing).

Project staff retrieved this file, converted it to a database, and developed programs to analyze the data. Gross statistics for the Archie FTP sites appear in table 3.

Table 3 FTP Hierarchy

Level	Total	Average
Sites	968	N/A
Directories*	169,718	202.77
Files	2,089,544	2,524
Size (no. of characters)	101,021,677,299	122,086,856

* "Directories" refers to the lowest node directories having files as their contents.

Analyzing the granularity of FTP sites provides another informative view of this information world. The structure of the FTP site itself provides additional information associated with the electronic files it contains, for example, the name of the site, the directory names and structure, and the file names. These findings are shown in table 4. The largest site at the time of our sampling was nic.funet.fi, which contained 5,625,701,955 bytes of information.

Table 4 Granularity of Largest Site

	Largest Site	All Sites
Number of Top Nodes	9	6.65 (avg.)
Maximum Depth/File	14	16
Minimum Depth/File	1	1
Average Depth	5.64	4.09
Number of Files	106,579	2,524 (avg.)

The site in question has nine top-level nodes. These are the highest-level access points to the information contained at the FTP site. On average, four descriptive elements precede the file name, although granularity at the directory level ranges from one (the file is at the highest level directory) to 16 (15 directory nodes precede the file name).

Further analysis among sites reveals additional information. For this study we randomly selected 20 sites to determine their size by file count and bytes. Findings for these 20 sites appear in table 5.

Table 5 Description of 20 FTP Sites

Name	No. Files	Bytes (100Mb)	Avg. (100,000)	STD	Maximum
a.cs.uiuc.edu	724	0.85	1.17	3.87	60.98
apple.com	11,119	4.36	0.39	2.24	86.47
boombox.micro.umn	172	0.23	1.31	3.30	20.48
cica.cica.indiana.edu	1,033	1.10	1.07	2.41	30.38
csam.lbl.gov	31	0.09	2.81	8.58	46.85
dsl.cis.upenn.edu	126	0.05	0.42	1.20	8.16
finsun.csc.fi	152	0.03	0.21	0.51	3.99
giza.ohio-state.edu	7,438	4.95	0.67	1.74	72.88
hubcap.clemson.edu	613	0.68	1.12	2.63	37.68
jhname.hcf.jhu.edu	12	0.00	0.09	0.14	0.45
merit.edu	849	0.79	0.93	2.18	26.77
lth.se	38,701	582.50	0.15	1.09	1.33
nic.mr.net	17	0.01	0.78	1.03	4.22
paul.rutgers.edu	5	0.01	2.37	2.60	6.03
research.att.com	105	0.88	8.36	30.26	293.83
shemp.cs.ucla.edu	51	0.09	1.67	3.36	16.96
suna.osc.edu	18	0.02	0.90	1.05	3.91
turbo.bio.net	300	0.17	0.55	1.53	19.79
uop.uop.edu	31	0.04	1.28	1.47	5.21
wategl.waterloo.edu	60	0.03	0.46	0.93	4.84

This analysis immediately reveals that FTP sites vary greatly in size, with a few very large sites accounting for a highly skewed distribution of information. Thirty percent of the sites account for 96% of the information, or 1.65 Gbytes of data.

FTP sites may provide additional information about their contents. Such information may be contained in files named "README," "Index," or some variation thereof. The 20 random sites analyzed contained 2,279 directories but only 239 "README" files. Only one site had a ratio of README files to directories greater than 1, and several sites had no informational files at all.

Characterization of File Contents

Project staff are experimenting with automated methods of categorizing electronic files based on information contained in path and file names. A database was constructed containing the complete path name for each file in the 968 FTP sites. The path names were stemmed and unique names collapsed to create a list. The list was scanned to determine which of the top 500 top directory names were most descriptive for the intended purpose, and from these, a dictionary was constructed. Files can now be processed automatically and divided into major groups. Results from processing a database of 20 randomly selected sites appear in figure 1.

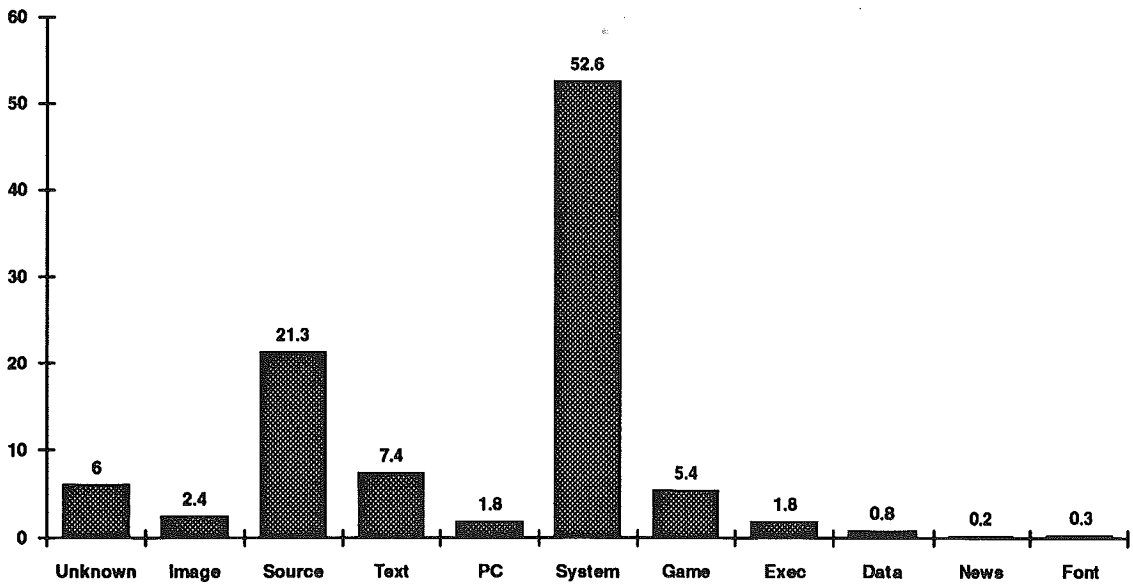


Figure 1 File Types by Percentage

The dictionary will be modified and expanded to include other bits of meaningful information such as file name extensions. The goal is to automate to the extent possible database creation and analyses, and to lay groundwork for future automated systems to assist the cataloging of electronic files in a wide area network environment.

NETWORKING AND TRANSFORMATION OF CATALOGING EXPERTISE

Paper submitted for
presentation and publication
at
the 1992 ASIS Mid-Year Meeting

by

Ling Hwey Jeng, Ph.D.

Assistant Professor
Graduate School of Library and Information Science
University of California, Los Angeles
405 Hilgard Ave.
Los Angeles, CA 90024-1520
(310)206-9366
BITNET: ICZQLHJ@UCLAMVS

NETWORKING AND TRANSFORMATION OF CATALOGING EXPERTISE

Ling Hwey Jeng, Ph.D.
Assistant Professor

Graduate School of Library and Information Science
University of California, Los Angeles
Los Angeles, CA 90024-1520

ABSTRACT

The development of expert systems for cataloging has had limited success to date. While there has been some promise in the area of advisory or training systems, no system has been able to achieve satisfactory performance in practice. To understand the lack of success and practicality in current cataloging expert systems, one must examine the process of building the knowledge base for such systems, especially the role of human expertise in it. This paper discusses the nature of human cataloging expertise and how it is networked and transformed in an institutional environment. A case study to investigate the network patterns and transformation of human cataloging expertise of the staff at the Cataloging Department of National Agricultural Library is described.

INTRODUCTION

Many have pointed out the potential of expert systems for research, education, and practice in library cataloging (e.g. Ercegovac, 1984). Ideally, for an expert system to perform intellectual work of cataloging, several components must be in place (among others, Hjerppe & Olander, 1989; Davies, n.d.): The system must be able to recognize bibliographic data electronically. The expert system must be able to interact with various bibliographic utilities and local online catalogs for screening and modifying copy cataloging records as well as updating holding information. The system must have access to authority files for verifying headings and maintaining authority records. In other words, the expert system must possess a knowledge base of enormous size that contains cataloging rules, term definitions, local policies and heuristics. Its inference engine must be able to learn from its own conduct and develop strategies to cope with more frequently encountered cataloging problems. Weibel et al. demonstrate in their study (1989) that it is desirable for the expert system to concentrate on items easily coped with, rather than processing a large amount of rules to deal with exceptions in cataloging. It must, in short, be able to distinguish between items it can process

and those needing human intervention.

Numerous experiments and prototype expert systems have been developed for cataloging purpose (e.g., Davies and James, 1983; Hjerppe et al. 1985). More recently, a prototype expert system for cataloging cartographic materials, MAPPER, was developed as a doctoral dissertation project at UCLA. MAPPER marks the first significant departure of approach to the development of cataloging expert systems, by demonstrating the first time that knowledge acquisition of human interpretation of cataloging concepts (i.e., map authorship) and of cataloging rules for cartographic materials can be systematically accomplished. The significance of this project is that, rather than concentrating on direct conversion of cataloging rules into its knowledge base as done in most other cataloging expert systems, the designer devoted much of her time to the analysis of human interpretation of map authorship and the statements of both intellectual and production responsibility in description (Ercegovac, 1990). It confirms the feasibility of developing expert systems in cataloging for the purposes of advising and interactive tutoring.

LIMITATIONS OF EXISTING CATALOGING EXPERT SYSTEMS

Most existing cataloging expert systems lack problem analysis and problem solving strategies. One example is the assumption made by many designers of cataloging expert systems on the adequacy of procedures and knowledge stated in cataloging rules such as AACR2 (Jeng, 1991). Another difficulty encountered by most existing cataloging expert systems is the inadequacy of assistance from human experts in cataloging. Most cataloging expert systems avoid dealing with human interpretation and heuristics (Hjerppe et al. 1985). MAPPER is the only attempt to overcome this difficulty with satisfactory results.

In addition to the limitations cited in most expert systems across disciplines (Liebowitz, 1989), cataloging expert systems face two major obstacles: (1) the enormous size of public knowledge (Davies, n.d.; Hjerppe et al., 1985) and (2) the extremely heavy emphasis on human interpretation (Hjerppe et al., 1985) as shown in numerous textbooks and training manuals using examples in learning. The lack of study on learning from examples in cataloging often leads researchers to conclude that cataloging is simply too complex or too arbitrary to be systematically codified into a structural knowledge base for an expert system.

EXPERTISE AND KNOWLEDGE BASE

It is apparent that the limitations of cataloging expert systems center around the understanding of expertise and its role in a knowledge base.

Expertise can be defined as a high degree of skills, dexterity or knowledge of a specific subject area. Johnson et al. (1988) describe expertise as a kind of operational knowledge. "It is characterized by generativity, or the ability to act in new situations, and by power, or the capacity to achieve problem solutions." Expertise differs from style of behavior in problem solving in that it is "a set of requirements that must be satisfied in order to solve problems in a given domain." (Johnson et al. 1988).

Glaser & Chi (1988) describe the characteristics of an expert as a person who works very well within a very specific domain but does not necessarily excel in other domains, who remembers a lot of things and has very high level memory structure, who solves problems by taking time to analyze the problem with strategies and reaches the solution quickly and accurately. Such a person also monitors his or her own performance and learns effectively from experience. Posner (1988) further points out that human experts spend less time learning how to cope and how to recognize, and more time analyzing problems. They can reproduce a scenario easily and quickly and have high level, long-term commitment for their domain (Larking & Simon, 1980).

A knowledge base consists of fact base and rule base (Richardson, 1989). The fact base, as described by Richardson, consists of declarative knowledge such as term definitions and standards. The rule base consists of procedural knowledge; i.e., standard problem solving strategies. Mussi & Morpurgo (1990) further point out that the rule base encompasses two kinds of knowledge: (a) strategic knowledge, which guide the diagnostic processor in deciding "which is the best action to execute next" when the problem of choosing among actions arises; and (b) knowledge about the entity under diagnosis, which includes the knowledge about functions and components of the entity, faulty parts and the most common misconception about the entity and how likely the particular faulty part can result in certain mistake.

An example of knowledge about an entity under diagnosis in descriptive cataloging is the identification of an alternative title. Rule 1.1B1 defines an alternative title as the second part of the title connected with the first part by the word or" (Anglo-American Cataloguing Rules,

1988). The rule prescribes related actions that a cataloger must take in transcribing such an alternative title; i.e., to transcribe the first part of the title proper, followed by a comma, a space and the word or and another comma and a space, and then the alternative title. In the case of a book on fundamentalism in biology, Science or Religion, the title serves as a title proper consisting of three words. An obvious faulty part in this title is the word or which can be easily interpreted, according to the AACR definition given in the rule, as the connecting word or for an alternative title. The proportion of all title propers which contain the word or that can be mistaken for having alternative titles will determine the likeliness of this particular faulty part to result in such a mistake.

In short, expertise means not only knowing lots of facts, but also knowing about strategies of dealing with problems and decisions. It involves not only the knowledge of what is right, but also the knowledge of what may go wrong and how to detect the faulty parts quickly.

CATALOGING EXPERTISE

Research on cataloging expertise has already begun in descriptive cataloging. Jeng (1987, 1989, 1991) paves the way to the understanding of human interpretation of bibliographic data in her empirical studies of the visual and linguistic cues of bibliographic data presented on about 200 title pages. Similar visual and linguistic cues were found important in identifying and interpreting bibliographic data on title pages in an experimental study conducted by Weibel et al. (1989) reporting a success rate of 75% in identifying and interpreting bibliographic data on title pages using visual and linguistic characteristics codified in only sixteen rules. Molto and Svenonius (1991) propose an algorithm for identifying corporate names by creating a machine-readable corporate name authority file and matching character string sequences on the title pages with those in authority file with high success rates.

Another aspect of research on cataloging expertise is the formation of public cataloging knowledge in cataloging presented in various rules and standards, such as AACR2. Codification of such public knowledge into the knowledge base of an expert system is essential as it serves as the basis on which human heuristics in implementation and interpretation of rules is added. In the study of the logical structure of such public rules in a knowledge base. Jeng (1991) argues that rules for description as they are presented and grouped in the mnemonic structure of Part I of AACR2 cannot be used as logical base for codification. The

rules must be further studied and broken down into logical condition/action pairs before they are codified into the knowledge base.

RESEARCH QUESTIONS

The above studies on title pages and on cataloging rules represent only a small portion in the realm of cataloging expertise. There remain areas of cataloging expertise yet to be tackled. A feasibility study for the Cataloging Expertise Project (Jeng, n.d.) was conducted to investigate the following two general questions: (1) What is cataloging expertise exactly? and (2) What is the scope of cataloging expertise? the study of differences between expert catalogers and novices in terms of the kinds of knowledge they possess and the level of competence involved. The latter involves the identification of components of expertise knowledge and performance of various interpretations and judgements.

The feasibility study attempts to define a conceptual model for the research methodology of such an investigation within a narrow subject and institutional domain; i.e., the cataloging activities at the National Agricultural Library (NAL). The model defines cataloging expertise as having five task categories and four skills facets. The five task categories are (1) searching bibliographic databases, (2) determining access points, (3) interpreting bibliographic data, rules and products, (4) identifying and prioritizing problems and special cases, and (5) administering performance. The four skill facets are (1) technology, (2) professional tools, (3) communication skills, and (4) subject specialty and foreign languages.

The remainder for this paper describes the on-going research and methodology on one particular issue in the Project; i.e., what network of cataloging expertise exists in the institution, and how expertise is transferred from the experts to novices within the institution.

DOMAIN SPECIFICATIONS

The domain of study in the Cataloging Expertise Project is limited to descriptive cataloging. This means the process of describing an item and the provision of name and title access points for the item. It deals with the processes of searching bibliographic databases; interpreting bibliographic data, rules, and products; description; determining main entry and added entries, and authority control. It does not include the aspects of subject

cataloging and indexing (such as in AGRICOLA).

SITE SPECIFICATIONS

The author has chosen the Cataloging Department of National Agricultural Library, situated at Beltsville, Maryland, as the study site. National Agriculture Library, since its creation, has been the authority of cataloging both research and popular materials in the area of agriculture. It is one of the leading national libraries contributing authoritative bibliographic records to large scale bibliographic utilities, such as OCLC. It also participates in numerous cooperative cataloging projects dealing with original cataloging and authority control, and in the development and implementation of cataloging standards. In addition, NAL has also been one of the leading libraries continuously engaged in research in various aspects of cataloging. NAL currently has a collection of 1.8 million volumes, mostly monographs and serials. Its Cataloging Department, led by Mrs. Idalia Acosta, consists of 17 professional catalogers and 16 cataloging technicians at various job ranking levels. Like most other cataloging institutions, a convenient distinction between the two groups is that the professional catalogers are experts and technicians are novices.

The library was chosen to be the study site for the following reasons: (1) It has a concentrated subject area, namely agriculture. The concentration of subject area provides librarians opportunities to establish their own subject specialty over time, and also provides librarians consistent parameters to acquire the subject specialty. (2) The library's active participation in national and regional level shared cataloging projects requires a high level of original cataloging and policy making activities among catalogers. The quality of such activities requires catalogers to have a high level cataloging expertise. (3) As a federal agency, NAL has formalized many procedures and knowledge through documentation such as job descriptions, policy manuals, and memoranda. The formalization of a knowledge base provides easier public access to the cataloging expertise involved. (4) The Cataloging Department employs more than 30 professional and paraprofessional catalogers. The staff size contributes to the objectivity of the project results.

DATA COLLECTION

Two unstructured small-group interviews were conducted with the Head of the Cataloging Department and three

professional catalogers in the Department. During the interviews information about the basic operation, workflow, and work patterns, procedures of recruitment, and performance evaluation of the Cataloging Department of NAL was gathered. Also gathered was documentation related to the organizational structure, job descriptions, evaluation forms and cataloging policies.

Contents from the interviews and documentation were analyzed to determine preferred methodology for further investigating the kinds of expertise network and the paths of expertise transfer existing within the institution.

NETWORKING PATTERNS OF EXPERTISE

Researchers in expert systems often distinguish between two groups of practitioners: experts and novices. The convenient differentiation between the two; i.e., professionals are experts and technicians are novices, does not hold in this project. Observations from interviews and documentation suggest that expertise can only be measured in relative terms. A technician in the Cataloging Department often has more cataloging expertise, acquired from long-term experience, than a professional librarian in another department with classroom cataloging knowledge. The technician certainly has more cataloging expertise than a non-librarian. In the monthly review session used for staff training and promotion, for example, the reviewers function as experts and reviewees, novices. On the job, supervisors are experts who review cataloging work done by subordinates who are considered novices. In each case, the distinction is relative, rather than absolute.

While it is not always possible to draw clear lines between experts and novices, because of the relativity of expertise, it is customary to describe expertise in terms of level of expertise in an institutional setting, such as NAL. Two levels of expertise existing in NAL are professional level and technician level. Each level is officially further divided into ranks according to federal job ranking system. Job descriptions for individual professional librarians and technicians collected during the interviews suggest that federal job ranking system actually serves as good indicators, in NAL, for the level of expertise of the individuals on the job. For example, the same duty, original cataloging, is performed by persons on all job ranks from GS-7 to GS-12 for material with different levels of difficulty:

GS-7	monographs, series, English translations ceased serials

GS-9	technical reports
GS-11	technical reports in French
GS-12	Germanic items and most difficult materials

TRANSFORMATION OF EXPERTISE

From private to public knowledge

The lowest job rank at which the transfer between private knowledge to public knowledge in NAL is done is a GS-11 position, in which case, a Technical Information Specialist with subject specialty is charged of formalizing the individual's private cataloging knowledge by "analyzing bibliographic databases to determine precedents in handling similar publications."

More often, transformation from private knowledge to public knowledge in cataloging, as found in NAL, is done at the job rank of GS-12. The transformation involves at least the following methods: (a) developing general policies for serials cataloging, (b) collecting and organizing information for revising cataloging procedures, and (c) analyzing problems related to cataloging procedures.

Private knowledge between individuals

The transformation of expertise between individuals takes on one of the following three formats: (a) the supervisor reviews cataloging work done by the subordinates occurring primarily between a professional at GS-12 level and another professional under his or her supervision. The review concentrates on the quality of cataloging record and correct application of cataloging rules. (b) monthly review sessions, in which some professional catalogers are designated as reviewers who check the cataloging records created by catalogers at lower-ranks; and (c) training sessions, in which the person at a higher job rank either gives training to newly recruits or acts as consultant or advisor to resolve discrepancies between records, conflicts between two other catalogers' interpretations, and to solve difficulty cataloging problems.

Transferability of expertise

As a whole, interpersonal transfer of expertise, in the case of NAL, is encouraged whenever possible. The Cataloging Department maintains a collegial atmosphere that facilitate such kind of transformation. The transformation of expertise between the Cataloging Department and other units within and outside NAL is not as obvious.

FURTHER RESEARCH

It is evident in the Cataloging Expertise Project that the network patterns of expertise in the Cataloging Department of NAL correspond closely with its federal job ranking system, and that the transformation of expertise from private to public knowledge and between individuals occurs as normal work relationships between colleagues at various job ranks.

One of the main components of the transformation of cataloging expertise is the process of cataloger's learning from examples. Although learning from examples is an effective way to train novices to be domain experts (Chi et al., 1989), the ability to link individual conditions and actions at the beginning of the learning process must be followed by a much higher level of knowledge organization, such as memory categories, or chunks (Halford & Wilson, 1980; Elio & Scharf, 1990). The high level thinking strategies on the learner's part to "generate explanations from examples that refine and expand the conditions and relate them to the solutions." (Chi et al, 1989) in the transformation process will be explored in the next phase of research in the Cataloging Expertise Project.

REFERENCES

Gorman, M. & Winkler, P. (Eds). (1988). Anglo-American Cataloguing Rules. 2nd ed. 1988 Revision. Chicago: American Library Association.

Chi, M.T.H. et al. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. Cognitive Science, 13, 145-182.

Davies, R. (n.d.). Expert Systems and Cataloging. In A. Morris. (Ed.). Expert systems in libraries and information centres.

Elio, R., & Scharf, P.B. (1990). Modeling novice-to-expert shifts in problem-solving strategy and knowledge organization. Cognitive Science, 14, 579-639.

Ercegovac, Z. (1984). Knowledge-Based Expert Systems: A Profile and Implications. IN Proceedings of the Fifth National Online Meeting (pp. 39-46).

------------. (1991). Design, implementation, and evaluation of an experimental cataloging advisor--MAPPER. Bulletin, SLA Geography & Map Division, (no. 163), 2-29.

Glaser, R., & Chi, M.T.H. (1988). Overview. In M.T.H. Chi, R. Glaser, & M.J. Farr. The Nature of Expertise. (pp. xv-xxviii). Hillsdale, NJ: Lawrence Erlbaum Associates Pub.

Halford, G.S., & Wilson, W.H. (1980). A Category Theory Approach to Cognitive Development. Cognitive Psychology, 12, 356-411.

Hjerppe, R., Olander, B., & Marklund, K. (1985). Project ESSCAPE - Expert Systems for Simple Choice of Access Points for Entries. IFLA 51st General Conference Linkoping: Linkoping University.

Hjerppe, R., & Olander, B. (1989). Cataloging and expert systems: AACR2 as a knowledge base. Journal of the American Society for Information Science, 40, 27-44.

Jeng, L.H. (1987). The Title Page as the Source of Information for Bibliographic Description: the Analysis of Its Visual and Linguistic Characteristics. Ph.D. Thesis. Austin, Tx.: University of Texas at Austin.

----------. (1988). The Language of a Title Page. In ASIS '88: The Proceedings of the 51st ASIS Annual Meeting. Atlanta, Georgia. Volume 25. (pp. 31-35). Medford, NJ: Learned Information.

----------. (1991). The Knowledge Representation of the Visual Image of a Title Page. Journal of the American Society for Information Science, 42, 99-109.

----------. (1991) The Structure of a Knowledge Base for Cataloging Rules. Information Processing & Management. 27, 97-110.

----------. (n.d.) Modeling Cataloging Expertise: A Feasibility Study. Manuscript to appear in Information Processing & Management.

Johnson, P.E., Zaulkernan, I, & Garber, S. (1988). Specification of expertise. In B.R. Gaines, & J.H. Boose (Eds.). Knowledge Acquisition for Knowledge-Based Systems. Volume 1. (pp. 125-145). London: Academic Press.

Larking, J., & Simon, H.A. (1980). Expert and Novice Performance in Solving Physics Problems. Science, 208, 1335-1342.

Liebowitz, J. (1989). How Much 'Artificial Stupidity' Do Expert Systems Possess? Information Age, 11, 225-228.

Molto, M., & Svenonius, E. (1991). Automatic recognition of

title page names. Information Processing & Management. 27, 83-95.

Mussi, S., & Morpurgo, R. (1990). Acquiring and representing strategic knowledge in the diagnosis domain. Expert Systems, 7, 157-165.

Posner, M.I. (1988). Introduction: What is it to be an Expert? In M.T.H. Chi, R. Glaser, & M.J. Farr. The Nature of Expertise. (pp. xxix-xxxvi). Hillsdale, NJ: Lawrence Erlbaum Associates Pub.

Richardson, J.V. (1989). Toward an Expert System for Reference Service: A Research Agenda for the 1990s. College & Research Libraries, 50, 231-248.

Svenonius, E., & Molto, M. (1990). Automatic derivation of name access points in cataloging. Journal of the American Society for Information Science, 41, 254-263.

Weibel, S., Oskins, M., & Vizine-Goetz, D. (1989). Automated Title Page Cataloging: A Feasibility Study. Information Processing & Management, 25, 187-203.

Electronic Publishing and Document Delivery
A Case Study of
Commercial Information Services
on the Internet

ASIS Meeting '92

Anthony Abbott
Senior Vice President
Meckler Publishing

I'd like to begin with a case study report of Meckler's publishing activity on the Internet and then go on to discuss some of the broader issues associated with commercial network publishing.

There is tremendous excitement now in the research, education, library and information management communities about the concepts and practices of **electronic networking**—i.e., the transfer of data among libraries, colleges, universities, research institutions, and commercial vendors linked together in a vast electronic web of individual sites.

Since its founding in 1971, Meckler Publishing has devoted its primary resources to the publication of books and periodicals and to the sponsorship of conferences and seminars whose purpose has been the exploration of new information technologies with particular emphasis on their utilization by librarians, information end-users and specialists, and the information industry as a whole.

As a publisher and a conference organizer we feel an imperative to be closely involved in this new stage in the evolution of electronic distribution.
Meckler's Print Program

Currently Meckler's print publication program comprises fourteen periodicals and between fifty and seventy books per year—including monographs, annuals, and other reference titles—on a variety of information technology topics.

Over the last year and a half we have started a newsletter on research and education networking, expanded our book program in networking topics, and begun publication of the first scholarly technical quarterly to treat electronic networking as a disctinct area of study.

But our markets extend beyond libraries. Such high-end technology publications as

Virtual Reality Report,
Multimedia Review,
Document Image Automation, and
HD (or High Definition) World Review

broaden the base of our readership and ensure a healthy penetration into the industrial sector where many of the newer technologies are born.

Conferences

We currently sponsor fifteen annual conferences in North America, England, and Europe, among them, Computers in Libraries, Computers in Libraries Canada, Computers in Libraries International, Virtual Reality, Virtual Reality East, HD World, and Document Imaging.

With our comprehensive emphasis on information technology and our coverage of electronic media, it was natural that we should seek to become involved in electronic networking as a publication medium.

The Internet

In the Spring of last year, recognizing that networking of information products and services was an area with potential major impact for scholarly publishing in general, we formed an association with the John von Neumann Center Computer Network located at Princeton University. JvNCnet, as the network is called, is one of the original networks comprising the NSFNET to provide Internet service to the Northeast.

JvNCnet is a separate network from the Princeton University campus network and with the bulk of its funding from the NSFNET coming to an end JvNCnet was charged with developing a commercial clientele for its networking capabilities.

We entered into an agreement with JvNCnet that would allow us to utilitze their Internet facilities for a number of purposes:

- to maintain an Internet email address (Meckler@jvnc.net);
- to use the JvNCnet file server to maintain a database of textual materials capable of being searched from a remote source; and
- to set up a mailing list for distribution of electronic data (i.e., electronic journal subscription list)

The database of Meckler documents was to be maintained in JvNCnet's Network Information Center On-Line or NICOL. NICOL is a user-friendly menu-driven application desgined to provide information about JvNCnet, its members, and other Internet activities.

It is based on the Princeton News Network system for UNIX. Internet access to NICOL is achieved by telnetting from a remote terminal to a specific Internet address chain, in this case: nicol.jvnc.net.

(Telnet is the standard remote log-in protocol for the Internet. It enables users to connect to any remote machine on the Internet direct from their terminal.)

In addition to our own file, NICOL maintains the Internet Resource Guide, the online public access catalog list developed and updated by Art St. George, and various other HELP documents.

Meckler's Internet Services

Our purpose as a scholarly publisher with Internet access was twofold:

1. Using what was essentially an open database, we saw the possibility of utilizing the network to direct potential readers to our full range of print publications and conferences.

2. As a publisher of periodicals and books and a sponsor of conferences, we saw the network as a way to extend that editorial role in a new medium.

MC2: Meckler's Electronic Information Service

Under the name MC2—an allusion to the initials of the company, Meckler Corporation—we began development of our Electronic Information Service. We can now see this development as being a two-phase program. I will discuss these two phases a little later.

Since we had been producing our publications internally for several years in a Macintosh-based desktop publishing environment, all of our essential documents—including periodical and book texts, promotional materials, and conference programs—were stored in electronic format.

We pulled a selected number of these files out of pagination software and reworked them into simple ASCII files, generated from MicroSoft Word. These documents were downloaded to JvNCnet and became our first public file.

For demonstration at the American Library Association's annual meeting in June 1991, we mounted this first test file, consisting of four documents:

- a general introduction to our proposed services;
- two topical indexes to our computer-related monthlies, *Computers in Libraries* and *CD-ROM Librarian;* and
- a full-text file of the May 1991 issue of *CD-ROM Librarian.*

This was an experimental file, with the full-text magazine issue mounted more as a test of the capabilities of the system than as a signal of our intention to mount an electronic version of that magazine on a monthly basis.

The indexes to two of our most popular periodicals, covering the last 6 years in the case of *Computers in Libraries*, and 5 years for *CD-ROM Librarian*, are also published in print versions.

Offering these two indexes online was an attempt to provide end-users broader electronic coverage of—and access to—articles appearing in these two periodicals for the time period prior to January 1991.

All four documents were keyword-searchable.

To facilitate use of each file, a string of options allowing the user to navigate through the document appeared at the bottom of the screen.

Following that demonstration, we mounted the following files:

- complete catalog of our books and periodicals;
- full conference programs for upcoming conferences; and
- order forms and registration forms.

In this way, we are providing what we feel is a complete electronic storefront to the full range of our publications and conferences.

Users searching these bibliographic and descriptive files can obtain information on any book, periodical, software program, CD-ROM, or videotape in our catalog, or any conference we sponsor, and can electronically place orders or register for conferences. They can also download order forms and registration information to use at a later time.The catalog and conference program files are updated as-needed.

Another editorial use that we saw for tapping the resources of the Internet was to mount questionnaires for response. Two databases of particular interest in this regard were *Dial In: An Annual Guide to Library Online Public Access Catalogs* and, for our annual directory *CD-ROMS IN PRINT*, a sub-section questionnaire on library-originated CD-ROM products. Questionnaires for both of these databases could be downloaded, completed, and returned either by mail, fax, or electronically to our email address.

Electronic Publishing Division

When we began to see the potential of our Internet involvement we established an Electronic Publishing Division. Under its auspices, several major projects would be coordinated.

One by-product of Internet access is the ability to send press releases and informational documents to various Internet and BITNET lists. This allowed us to make our electronic services known in a quicker and more pertinent manner than print publication (although we also undertook to print advertisements of MC^2 and continue to do so with increasing frequency). We were thus able to announce the formation of the Electronic Publishing Division, as well as each new enhancement to MC^2 as it occured.

MeckJournal

Once we had established the parameters of the bibliographic portion of the file—the catalog, the conference programs, order and registration forms—we began our editorial constructions.

Two major projects were undertaken in the summer with the intention of broadening our document availability services. The first was the creation of an electronic journal to publish in Internet-accessible form selected articles and other texts from our print library of

information technology periodicals. The second was to provide the foundation for our document delivery services.

MeckJournal was inaugurated as a monthly in September 1991. To date, the first three issues are browsable in MC^2 along with our other files. Current subscription to the journal is free.

The first issues have featured an introductory editorial, a news and comment section, and full text of from one to two articles taken from Meckler print publications.

While we were aware that mounting the journal on the NICOL database would give it a more formal and archivally safe status, as well as being able to draw in potentially more users, we also wanted *MeckJourna*l to have the forward presence of other electronic journals, such as *E-Journal, The Public Access Computer Systems Review, The Electronic Journal of Communication*, and *Psycoloquy*, among others.

We also wanted the journal to have as high a profile as possible within the electronic library community.

So, when our first issue was nearing completion, we posted a message to a number of library computer conferences and discussion lists whose readers would theoretically be interested in *MeckJournal.*

The response to this posting was immediate and exhilarating: within 24 hours we had over 250 respondents requesting to subscribe to our first issue. By the time we had assembled all the addresses and forwarded them to JvNCnet for development into an email subscription list, the total was nearing 900 subscribers. This number is growing constantly as word of the journal's existence spreads.

Shortly after this promotional posting we were contacted by the computing center administrator at one large eastern university requesting permission to load the journal on their campus-wide information network. We were then approached with a request to allow *MeckJournal* to be cross-posted to CompuServe's "Telecom Issues Forum." We allow and endorse both of these extended uses of the journal.

The ability to make *MeckJournal* available by electronic subscription also meant that we could reach a wider audience. BITNET users, unless their institution has an Internet account, cannot use the telnet function necessary to access our MC^2 files on JvNCnet's NICOL.

The added benefit of having a subscription list—distinct from a browsable database whose use was not monitored—meant that we knew who was interested, that we could play a more active publishing role, and that our subscribers could also communicate with us directly relative to the journal, and thus provide the grounding for a more interactive publishing environment.

Users wishing to subscribe to *MeckJournal* were instructed to send an electronic mail

message to our Internet address—Meckler@jvnc.net. The message reads: subscribe MeckJournal First name Last name full email address.

To make *MeckJournal* truly responsive to the electronic environment, we have established a rotating editorship among Meckler's print journal editors and have introduced a changing line-up of guest-editors. The selectivity that we have in choosing the contents of this publication, taking from here and there among the entire range of our resources, to us exemplifies the essence of electronic publishing's fluidity and immediacy.

MeckJournal: Communications Problems and Other Growing Pains

We were made aware by our responsive subscribership that the e-mailing of issue # 2 to subscribers in mid-November was accompanied by a number of problems. One problem -- as best as we can reconstruct it -- centered around messages sent to the publication by some subscribers who used a slightly wrong (but perhaps intuitively natural) email address, using MeckJournal in the command, instead of the correct address, Meckler@jvnc.net.
The "MeckJournal" part of the address kicked in the subscription command, so that comments meant to be received only by the journal's editorial staff were being sent instead to the entire subscription list. It was a day or two before this was discovered, causing receipt of some dozens of messages by subscribers and quite a lot of confusion and irritation.

This was a wholly unexpected incident. What we did was to structure our issue mailings so that the subscription command is activated only for the monthly e-mailings and disabled between times. If subscribers attempt to send a message to the subscription address, they will receive a "Host Unknown" or "Address Unknown" response.

It was also brought to our attention that the second issue appeared onscreen in a double-spaced format. It was not our intention to do this, and the issue itself, as it appears in our files is fully single-spaced. We have not been able to explain how the extra line spaces were introduced; perhaps in our email delivery from our editorial offices in Westport, Connecticut, to JvNCnet's office in Princeton; perhaps somewhere in the uploading procedure. In any case, we have begun to monitor this aspect of issue preparation so that the obvious electronic inconvenience of a double-spaced transmission does not recur.

It is also interesting -- in a trial-by-error sort of way -- to note that the first uploading of MeckJournal #2 to the NICOL database, had a problem with the text reading off the screen on the right-hand side. This was do to the fact, later discovered, that the issue had been prepared in 10pt type in MicroSoft Word, and then sent in a hard-return line-ending version of Ascii to JvNCnet. The number of characters per line, however, exceeded that readable by most screens, causing the lopped-off text. Luckily, this problem was discovered before the subscriber version had been mailed.

Tables of Contents Database—Periodicals

Our second major undertaking in Phase One was the creation of a database incorporating

the tables of contents from every issue of each of our periodicals beginning in January 1991 (and in some cases earlier).

This database would serve a number of purposes, but its primary use in the context of our electronic services would be to enable us to offer document delivery—by mail or fax—for any article, editorial, column, or news item that has appeared in any of our periodicals since the beginning of the year.

Our current periodicals include:

Academic and Library Computing
CD-ROM Librarian
Computers in Libraries
Database Searcher
Document Image Automation
Document Image Automation Update
Electronic Networking
HD World Review
Library Computer Systems and Equipment Review
Library Software Review
Multimedia Review
OCLC Micro
Research and Education Networking
Virtual Reality Report.

These fourteen periodicals—soon to be increased by two more—represent the range of our commitment to information technology topics.

The table of contents database now comprises some 1500 records covering the period January to December 1991. The online service will be updated quarterly. If demand for earlier documents is apparent, retrospective data entry of tables of contents of periodicals published prior to January 1991 will be effected and added to the database.

Tables of Contents Database—Books

A third project now in its initial stages, but projected for completion in the Spring of this year, covers the tables of contents of all the monographs in our book publishing program. This service will offer access down to the individual chapter level, with the aim of providing full customized document delivery and fulfillment.

The combination of both databases—periodical and book—will permit comprehensive access to a full range of individual texts and may overcome some of the expressed limitations of traditional monographic- and issue-based entities in favor of a more end-user oriented custom-publishing approach.

Recent books include:

Advances in Library Resource Sharing

Public Access CD-ROMs in Libraries: Case Studies
CD-ROM Local Area Networks: A User's Guide
Virtual Reality: Theory, Practice, and Promise
Mass Storage Systems
Essential Guide to dBase IV in Libraries
Case Studies of Optical Storage Applications
Library Technology for Visually and Physically Impaired Patrons

Contents of books will be listed comprehensively in this program and updated on a bimonthly basis. Again, the Internet user is able to browse any of these bibliographical listings without charge. If the user finds an item of interest, he or she may then electronically request delivery of the document—either a periodical article or a book chapter.

MeckFAX Document Delivery Service

The periodical and book tables of contents databases serve as the basis for our MeckFAX document delivery service. Any article, column, editorial or book chapter that has appeared in any of our periodicals or books can be ordered electronically, and delivered via fax or mail. By using the MeckFAX order form in the MC2 main menu, users can order documents for a straight $15 prepaid per document, specifying either mail or fax delivery. All orders will be faxed or mailed within 48 hours of receipt.

Eventually, as demand for electronic delivery is shown, delivery will include electronic formats.

Online Databases

Mounting the text of our major databases electronically is also an option.
Though we have only just begun to ponder the economic structure needed to ensure that vital print revenues are not diminished by electronic access, among the options available to us are charging either a flat "purchase price" against annual use of the database or evolution into a per-use fee structure similar to what commercial online search services have established.

Examples of the latter kind of pay structure include those for Mead Data Central's Lexis/ Nexis database and OCLC's EPIC service, both of which services are accessible on the Internet and are charged out on a per-use or connect-hour basis.Two databases we are considering mounting in this way are *Dial In: An Annual Guide to Library Online Public Access Catalogs* and *CD-ROMS IN PRINT.*

Over the next few months, the resources of our Electronic Publishing Division will be spent developing, refining, and updating these files and several new projects that are currently being researched.

Commercial Electronic Publishing

It would be appropriate here to address some of the broader issues of commercial network publishing.

If the concepts and beginning practices of computer-based publishing have done anything, they have forced us to formulate in a more streamlined statement than we have felt the need for before, exactly what our role as a publisher is in an era of electronic information. I would characterize that role, in its simplest form, as the creation and delivery of documents.

Let me define the key terms in that equation. "Creation", in its most complete sense, is taken here to mean origination— through the agency of authors, editors, or computer-based data manipulation.

"Documents" can be defined as any sort of data—textual, numeric, graphic, software, audio, or video—and be presented in any form or media—print, microform, audiotape, videotape, or electronic—that is wanted or needed by a specific consumer group..

And "delivery" refers to the transfer from the creator to the user by any and all means that (a) serve the user's convenience, need, and timeliness, and (b) allows the creator, as a commercial organization, to recoup creation and delivery costs, and to profit in such a manner as to permit it to continue in business as an information creator and provider.

With this streamlined formula in mind, some pertinent issues present themselves to the publishing community. I would like to note a few of these now.

Changes in the Research Process

The revolution in electronic publishing over the last few years has led to the development of a number of alternatives for the researcher that are challenging the traditional view of where information is obtained.

The electronic transfer of research documents, bibliographic files, and other data directly to the researcher's remote workstation, is shifting the time-honored dependency of the researcher away from the library to campus-wide information systems, online public access catalogs, and networked information resources.

The researcher has ever more increasing power in the shaping of information content and delivery. Moreover, the relative ease and cost-free or cost-hidden nature of electronic publishing, which can now be seen as the natural extension of the desktop publishing revolution, allows the individual scholar or group of scholars to present in an inexpensive and viable format a medium for discussion of ideas, publication of work in progress, and intercommunication, without the commercial mandate.

Changes in Publishing

The term "Custom Publishing" is one that takes its power from the fact that it is consumer-based. The end-user with less and less time to wade through complete documents requires access to those parts of books and journals that are appropriate to the research at hand. This sort of orientation is inherent in our—and other publishers' and agencies'—commitment to providing access to and delivery of individual chapters and articles.

On the one hand, modular publishing can be seen as a new revenue source, on the other it poses concerns about the dropoff of traditional book and subscription purchasing. For the time being, sale of chapters and articles on an individual basis will likely be an ancillary source of income to publishers. One is comforted by thoughts of the possessive nature of scholarship—the scholar's desire to have everything in paper and to have one's own copy will work to generate an additional area of sales revenue, at least at first.

In an article in a recent issue of *The Public Access Compuer Systems Review*, an electronic journal edited by Charles Bailey, Ann Okerson noted with precision the parallel modes of publication of a text that electronic media make possible:

She stated that to maintain the necessary levels of revenue the same article could be published in the following ways:

- in a paper journal;
- in single article delivery (by mail, fax, or email to purchaser);
- in a compendium of articles from several different journals;
- in a collection prepared to an end-user profile;
- in a publication on demand structure; and, finally,
- in a networked delivery to research facilities and institutions.

By maintaining a fluidity in our own internal publishing environment and through our network connection, we feel that we can now or will soon be able to provide these various formats with relative ease.

Bibliographic Control

Libraries are in the business of bibliographic control and bear some of the responsibility for archiving and making accessible publications created and delivered in electronic media. To my knowledge, there is little formal literature on managing collections of electronic journals in libraries or on making these invisible documents available to the library clientele at large. Information management is one of the most significant challanges that libraries face in this new stage of publishing history.

Another barrier to fuller acceptance of discrete electronic publications such as electronic journals is the fact that indexing and abstracting services, traditionally the gateways to document access, have yet to absorb the electronic journal in a normal way into their activities. It is instructive to learn, however, that the joint project of OCLC and the American Association for the Advancement of Science, *The Online Journal of Current Clinical Trials*, due to begin publication in April, will be indexed and abstracted from its first issue onwards in the BIOSIS database.

The Electronic Journal of Communication/La Revue Electronique de Communicatin, jointly produced at the University of Windsor and the University of Montreal is also actively involved in seeing that its contents are included in standard abstracting and indexing services.

The great publicity given both these and similar individual efforts will undoubtedly raise

the level of awareness of electronic media and forward the discussion of the purchase and management of such publications in the library.

Commercialization

Under the term "Commercialization" come a number of subtopics. Among them—fee-based network publishing, document delivery, advertising on the networks, and copyright protection.

With the regional networks forced to battle for both non-profit and commercial members and governmental funding of the proposed NREN comprising little more than seed money, it seems inevitable that the networks, while retaining their essential base in the research and education community, will succumb to the market economy that all transactive societies appear to evolve to.

One is interested to see the moderated online information distribution service that EDUCOM is mounting for the benefit of its Corporate Associates. Called *CAPNEWS* (after Corporate Associate Program), the service will distribute to subscribers via BITNET, information on new products and services provided by EDUCOM's Corporate members. It will essentially be a catalog of new products, with, presumably, the ability to order them.

Pricing

Meckler is not at the moment charging directly for any of the services we are offering on the Internet. Our plans—over the next year—include establishment of a fee-based subscription service on the network. Whether this will be an enhanced version of *MeckJournal* or an electronic subscription program coverng the whole range of our print periodicals, has yet to be finalized.

Certainly, we have the technical ability and the Internet connection to permit us to offer our publications in electronic format. The question of how to price this category of subscription items is a challenge we are beginning to face as we work to bring a profit to our current electronic publishing activities.

Right now, the single method of publishing which seems to provide some safeguard against revenue loss due to competition by electronic access, is through development of an email subscription list. In this manner, the publisher has control over dissemination of a publication and can deliver based on a prior payment.

Copyright and Intellectual Property

Copyright protection and revenue stabilization are inseparable in the networked environment. Electronic infringement of copyright is identical to revenue loss: these two concepts should be fused in the minds of publishers and their authors.

While there is no guard against any subscriber misusing the data after it has been e-mailed to an account—loading it onto a local area network, bulletin board, or campus-wide information system—the subscriber's possible membership in a copyright protec-

tion agency may have some effect in keeping violations of both the letter and the spirit of traditional copyright restrictions to a minimum.

An alternative would be to grant site licenses in addition to traditional subscription fees. Payment of the license would allow a limited or unlimited amount of copying and distribution.

A further method of diminishing revenue loss and safeguarding rights would be to tie in an electronic subscription and site license package with purchase of a print subscription, perhaps at a reduced rate—I described this earlier. The traditional revenue base is protected but the benefits of electronic access—including, importantly, dissemination of an author's work—are forwarded.

Copyright has always been a problematic issue, one that is subject to much discussion and which has spawned a huge industry and an extensive literature of its own. There are those who suggest that existing copyright regulations could be adapted to safeguarding the rights of authors and publishers in a networked environment.

Some have called for the establishment of a publisher agent in the network milieu; perhaps an existing agency like the Copyright Clearance Center could be configured to handle the new issues.

Some have argued that a new coalition of publishers' representatives would need to be structured. This is a question only more network experience by all relevant parties will be able to answer.

Perhaps a new body to which the library and individual and corporate copiers would apply as members will evolve. Use and copying fees could be collected through notification of use, as now occurs with CCC's programs in the realm of paper copying.

Document Delivery

With the availability of all of our documents in electronic form, our database of tables of contents becomes an ordering mechanism for document delivery.While there are a number of established independent document delivery services already in existence, the publishers who stand the greatest chance of profiting the most are those who have the power to provide delivery services themselves.

The delivery of documents will become a greater issue than in the past in the way a publisher operates and will have profound influence—perhaps in the long term—on the survival of publishers into the electronic era. For it will be one of the publisher's great assets to control the promotion of and purchase price for the documents it creates.

In a sense many publishers have done this for a long time. Offering reprints of articles from their publications is a standard source of revenue for many publishers.

With the network gateway to this service, publishers will have an additional method of promoting their traditional document delivery services, but they will also be able to work

their array of resources into the kinds of publications wanted by the end-user.

Already, I understand, there are software programs being developed to allow computer-mediated requests for documents by number, providing very fast credit card validation and associated fax delivery of document and purchase receipt.

Advertising

There is concern among commercial publishers that advertisers will not endorse electronic efforts, that if an advertising-bearing periodical publication is converted to electronic format advertising will fall off and thus a traditional source of revenue for many publishers will diminish.

Advertisers will come to electronic publishing with their own models of how to advertise in a digital medium. If multimedia technology and network bandwidth research progress with the swiftness that current assessments indicate, before long it will be commonplace to receive good graphic (as well as audio) representation on the end-user terminal, in a form that is not, we would hope, obnoxious to that user. That is not the problem.

The difficulty will be the traditional one of gaining the market. If it can be proved that the end-users are there, needing the publication, then advertisers will be among the first to demand access to that market.

Right now, however, the case cannot be made that the electronic audience is a broad and strong one. There is still an appearance of fringism, of superfluity in electronic publishing activity.

This appearance will change, and change quickly, over the next twelve to eighteen months, fueled by such higher-profile and higher-budget efforts as those of organizations like OCLC, EDUCOM, Faxon, and CARL.

One of things that we are doing, as a publisher of print products to the library community, is to emphasize the concepts of electronic publishing—in articles, columns, advertisements—and in our own way link the electronic journal to the other resources that libraries and information handlers must come to manage.

We are developing a monthly column authored by Erik Jul of OCLC, which will appear in *Computers in Libraries* by the end of the year,on the topic of electronic journals. In addition, we are developing an annual review of advances in electronic publishing.

One of the missions of this column and this annual will be to bring to the attention of the library community on a continuing basis the wealth that electronic journals offer. Issues such as collection management, standards, bibliographic control, archiving, abstracting and indexing, and other concerns about the medium will be addressed.

We hope that this effort—along with our others—will grow into an important forum for the treatment of these issues and that librarians and publishers will contribute their views on how to bring networked-based information to a wider audience.

One Final Word

As much as editors, publishers, and librarians are intrigued by these technological advances in information dissemination, the profitability in the transformation of print to electronic media and in the creation of wholly new electronic documents will only occur at the insistence of the end-user—the same end-user who will demand of publishers and librarians alike the power to create publications profiled to specific research needs.

For further information:

Meckler Publishing
11 Ferry Lane West
Westport, CT 06880
(203) 226-6967
Fax (203) 454-5840
Meckler@jvnc.net

Alan M. Meckler, President
Anthony Abbott, Senior Vice President
Nancy Melin Nelson, Vice President, Information and Technology

---

To access the MC(2) file on JvNCnet, users with VT100 terminal emulation can telnet to nicol.jvnc.net and type nicol at the login prompt. Select MC(2) from main menu. No password is needed. To subscribe to MeckJournal (from the Internet or BITNET), email to Meckler@jvnc.net with the message, subscribe MeckJournal First name Last name full email address.

---

presentation for American Society for Information Science
1992 ASIS Mid-Year Meeting
May 27-30, 1992
Albuquerque Hilton
Albuquerque, NM

A Cooperative File-Library for Electronic Experimental Research:

The Consortium for Lexical Research

SPECIAL REPORT

Yorick Wilks and Sylvia Candelaria de Ram

Computing Research Lab
New Mexico State University
Las Cruces, NM 88003
yorick@nmsu.edu OR yorick at nmsu [bitnet]
sylvia@nmsu.edu OR sylvia at nmsu [bitnet]
(505) / 646-5466 sec.; 646-6520; fax: 646-6218

ABSTRACT:

This presentation does not describe research, but reports the formation and sketches the nature of a facility being formed for doing research. It is not a description of an automated technical library, but of a cooperative archive of electronic materials for specialized research. The software and documentation products of research being gathered into the archives from and for Consortium members are tools for further research.

Contents and access. *As an electronic text facility, the materials of the library are not two-pound books carried in the hand to be returned in some days. They are volatile packets of electronic information to be picked up by a copy process, some small, some enormous, some legible, some encrypted. The information may be compressed and then have to be uncompressed. Some packets may run as programs in a given operating system and language environment, with or without further compilation; some may be already compiled. Others may be texts rather than programs. Some of the pickups may occur without personal transaction of the keepers of the archives. Contributions too may be provided in various ways without presence of the agents of the exchange at the same site and at the same time. Other pickups may entail exchange of concrete objects such as magnetic diskettes or tapes, or fees or paper documents.*

Consequences of electronic exchange. *The goal of the formation of the Consortium is to foment Lexical Research. Just as a traditional library functions to inform its would-be inventors of "the wheel", a major goal of the Consortium archives is to permit faster progress in computer-based research involving* ***vocabularies****. Now traditional libraries have been a critical spur for development of knowledge by making it easier for people to build upon what is already known about. This new, very fast medium of communication of "information patterns" goes beyond permitting literates to access information. It permits computer-literates with the proper equipment to get their computers to do something that someone else has figured out how to do, without having to understand themselves how it is actually being done. Instead of "re-inventing the verb-wheel", as it were, researchers can feed their computers what it takes to actualize a verb-wheel on the spot. Then they can use their time for putting that wheel in place as a cog in a system for processing natural*

language that is much larger and more complex than any they would be able built up alone from scratch.

Consortium members all over the world share an aspiration of great importance: to achieve processors adequate to tasks of using language like we do, processors to assist us with those tasks right away on a scale matching the burgeoning scope of our current busy collection and communication of information across time and place. Pooling resources electronically is clearly needed, and the Consortium for Lexical Research is an assay in how that may be done. Contributions are welcome.

Research — a driving force.

Research is driving academia and industry today, and hence government as well. Research means change, new things to be taught, bought, and integrated into society. Seeing to the needs of research is therefore of paramount importance these days.

Research is specific — into how this phenomenon works, how to build some particular thing, or what the answer to questions arising from a particular vantage point must be. The resources to work out these things are of specific kinds too, even when exactly what will lead to the solutions sought is not known a priori. Existence of the resources is not enough. Researchers must have access to them. This is a bit paradoxical, because exactly what researchers need or can make do with is not always clear, even though without that whatever-it-is the discoveries or inventions cannot be accomplished.

Research into how to bring computers *effectively* into our communication and information supply loops is presently one of our hottest and most urgent endeavors. For the world is not shrinking, but being tied together. Getting things and masses of information across can be faster, with distance less of a barrier. But that's when the system works right! Designing it right to do these tasks that never were possible before is going to take comprehension of how communication works, and tremendous linguistic resources, for languages are enormous and dynamic. We have encoded in our many languages all kinds of social and material history, and set out innumerable "recipes" for discriminating, careful thought: What is like what (synonymous or antonymous), systems for analysis (special vocabularies), and even social volition (argument structures).

Berries for berry patches.

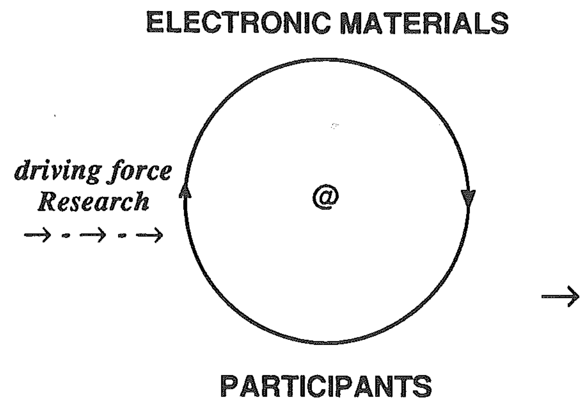
Gathering these together is one of the first jobs in getting research in computer language processing[1] what it has to have. To this end, researchers are anxious to share the language resources they have compiled so far — lexicons and tools for using them, such as parsers — so they can get on with the communication part of it, rather than trying to gather up words and start from scratch in every project by laboriously labelling some of each word's properties. There are too many words, and too many usages to make that practical! [See Figure 1.]

CLR at CRL: A hub for information conservation and exchange.

A Consortium for Lexical Research serves as a focus for this activity. Labs and individuals all over the world who are engaged in lexical research are involved. The work is sponsored by the Association for Computational Linguistics, DARPA, and the Computing Research Lab of New Mexico State University directed by Yorick Wilks. The archives of the Consortium, the CLR, are sited at NMSU's CRL. This is a node in the worldwide electronic communication network (internet) and affords up to date computing facilities. The lab is a center of research in computer language processing (often called NLP in the trade for Natural Language Processing) and electronic dictionary research. The materials stored in the archives are in electronic form. They are available instantaneously to Consortium members, be they in New Mexico or in New Zealand or even in New York or Tokyo or Torino. There

[1] Often referred to in the trade as NLP or Natural Language Processing.

*** RESOURCE CYCLE ***



EXCHANGE PARTICIPANTS:

RESOURCE PROVIDERS
&
RESOURCE USERS
@
CLR
—
Consortium for Lexical Research

Figure 1. Exchange of electronic information to get language research really rolling. Set at the hub is the Consortium for Lexical Research, which receives and distributes lexical materials of all kinds for use all over the world in building up a cross-linguistic communications base.

are of course no book cards for patrons to sign. There is no filling out and mailing in of a form for each item brought out as in some consortia set up under electronic tape distribution procedures. (An example is the Inter-university Consortium for Political & Social Research, dating from 1962.) And there is an unlimited number of copies available using the favored procedure of online file transfer (ftp, file transfer protocol). Depositing is just as easy, and uses the same protocol (`ftp ... put FILE` instead of `ftp ... get FILE`).

Going by the book and going past the book: Archive Administration.

It is, of course, more complicated than that. Members have to become members. Some materials are under copyright and restricted to research use. Some bear fees, and are being made available for research and product development where their use will have to be under contract for further distribution. These *heavily encumbered* materials are distributed under special accounts in encrypted form after the signing of paper agreements. They are administered differently from *lightly encumbered* and

unencumbered and public domain materials. Still, this is not a paper library, and there are many new potentials in its use and organization. [See Figure 2.]

"Multiplicated" materials

Long-term technology advance:
•→○

PALPABLE C O P Y • based on paper-and-ink

•traditional libraries	***•→•direct loan, return***
inter-library loan (ILL)	***indirect loan, ± return***
film (microfiche etc.)	***visual image loan, transfer to paper***
⊙electronic tape archives	***⊙ ⊙physical duplicate of electronic image***
○ electronic communication archives "e-archives" [e-access]	***electronic transfer ○virtual copy***

VIRTUAL C O P Y ○ based on polarized pulses, ± active

Figure 2. Our day is seeing a gradual evolution of the library from one based on paper-and-ink to one based on electronic encoding. The palpable and visible book [•] is joined by the impalpable but computer-readable virtual copy [○]. New provisions are required for copyright and use restrictions to maintain the traditional library-patron-author balance (see Candelaria de Ram 1991[2]).

Foremost is the fact that what is "checked out" of the archives may never be read by the Consortium member at all. The researcher may not want to read it, in fact, if it is a program in binary code and practically illegible on that account. The Consortium member may simply run it, and use the program to do something like find a set of synonyms or scan a text for them or provide a user interface.

Going by the book and going past the book: Running while Reading.

Using the archives requires a certain amount of computer literacy as well as equipment. The facilities operate with an operating system in wide use in the academic computer research community, which

[2] Candelaria de Ram, Sylvia. 1991. "The Consortium for Lexical Research", p. 117-119 in the proceedings, Workshop on Language and Information Processing. 'Systems understanding people, people understanding systems', Proceedings, Conference of the American Society of Information Sciences (ASIS). Washington, D.C. Oct. 27-31.

is under continuous co-operative development at many sites. In this Unix operating system, electronic code in many encodings can readily be stored and accessed over net or with modem connections. Encodings aplenty have to be translated from one to the other and a portion of the archives consists of programs for members to do just that. Brought in with software into the archives is *documentation on how to run the software*. There are several kinds. — There is inline documentation in the code. There are instructions. There may be explanatory material, elaborating the vantage point of the originators, their analysis framework, *identifying the linguistic corpus on which the materials are based,* and giving examples. Depositing all of these is encouraged, for this is after all a research facility for the research community and this information is extremely important in interpreting and building up our understandings of what our lexica are and how they work. How can a researcher confidently use something built by someone else if it is not clear what it is and what it does? Good documentation is an essential for the resource-sharing enterprise. That should hardly be a surprise to computer people, to lexicographers, or to librarians.

The archives themselves bear several kinds of documentation. [See Figure 3 and Appendix.] Several catalogs of materials are written up for members' use. There is the *short-catalog*, with its brief naming of the material and its originator. It includes the specific file names for getting to the proper part of the archive directories to copy the selected files to the computer memory at the member's site. The *paragraph catalog* is comprised of paragraph-length blurbs describing each set of materials to help researchers get an idea of what there is that they might find uses for. More detailed *information files* provide specifics about what hardware and software environments are involved, about copyright and use conditions, and may show examples of how the input and output look, etc. These are generally built from the originators' documentation and experience with the materials as the archive staff works with them. There are instruction files about how to use the archives when you first enter the ftp site by calling it in. Members also get information by email (and hardcopy mail) about what is coming in, and what is going on. A CLR Workshop was held in January 1992 to which key researchers and publishers in lexicons pooled their know-how to assure the success of the endeavor.

In addition to software and lexica in many languages and from many sources, portions of the archives are designed to provide for updates on what has been developed and how well it works, bugs that have been found, on what is going on at different research sites, and on what is available in the literature. By handling the materials and gathering feedback on them for their providers, the Consortium can streamline their development. In fact, pooling resources in this way promises to help smooth communication all over the world, by contributing toward one of today's major aspirations: Getting computers into shape to participate constructively in communication on a large scale. Contributions are welcome.

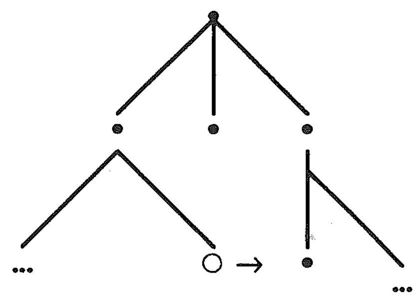
CONSORTIUM FOR LEXICAL RESEARCH

ACL Association for Computational Linguistics

Computing Research Laboratory (CRL)
NEW MEXICO STATE UNIVERSITY
Box 30001/3CRL/Las Cruces, New Mexico 88003
Telephone: (505) 646-5466/6520
Fax: (505) 646-6218
Email: lexical@nmsu.edu OR lexical at nmsu [bitnet]

** INFORMATION RESOURCE MATERIAL **

PALPABLE	VIRTUAL
read by humans	read by human-computer-complexes
visible	volatile
shelved materials	archived files
subject cataloguing	interlinked directories



card-index	computer-editor searches
materials borrowed	materials cloned □□
same returned *t*	derivatives deposited □ *t*

Examples of derived resources:

/ < / sub-lexicon (e.g., technical)

/ & / new-version parser in updated code

/ + / augmented lexical entries

/ @ / bug reports

/ → / indexed texts/terms

/ ~ / parallel in other language

Figure 3. The Consortium for Lexical Research (CLR) is not an automated technical library but a co-operative archive of electronic materials for specialized research of great import on society. It is based on the new potential for "multiplicating" electronic information to make it quickly available across long distances.

APPENDIX:
Sample catalog entries for archives of Consortium for Lexical Research

Sample entry from Consortium for Lexical Research short-catalog:

```
Doc File:          readme
Ftp File:          arcrexx.exe
Short Description: Self-extracting MSDOS REXX code for ARCSGML parser
                    (parser for getting tagged material from texts previously
                    tagged for Part-of-speech etc. with SGML markup system)
Ftp File:          arcsgmlc.exe
Short Description: Self-extracting MSDOS C source for the SGMLUG
                    ARCSGML parser
Ftp File:          arcsgmlh.exe
Short Description: Self-extracting MSDOS C header files for the SGMLUG
                    ARCSGML parser
Ftp File:          arcsgmlu-1.0.tar.Z
Short Description: A version of the SGMLUG ARCSGML parser fixed for Unix by J. Clark
Ftp File:          arctest.exe
Short Description: Self-extracting SGMLUG ARCSGML parser test files
Ftp File:          arcvm2.exe
Short Description: Self-extracting SGMLUG ARCSGML markup validator

Ftp Directory:     pub/tools/arcsgml/
More Info:         info/5
```

Sample blurb from paragraph-length catalog of Consortium for Lexical Research online archives, for same materials above.

ARCSGML is a set of tools for setting up and working with text that is tagged with your own specialized tags in SGML format. The tags permit you to label text structures (such as Part-of-speech, syntactic, morphological, semantic, or discourse structures). The parser is for selectively pulling out corresponding tagged pieces of text. The ARCSGML toolkit is for use in developing conforming SGML parsers, systems, and applications. A validator (for checking your tagged text) is supplied. It supports the standard SGML reference concrete syntax (beginning from 1983) in all features except LINK, CONCUR, and SUBDOC (although some hooks are in place to get you started on these). [The package was originally written to validate the 1983 working draft of the SGML standard, and was subsequently maintained to track the standard through its final phases of development, culminating in the amendment.]

Executable sourcecode programs for versions for PC and Unix C (MSDOS REXX, MSDOS C, and Unix C) are provided.

Sample information file for a tool for lexical research in Non-Roman contexts from the archives of the Consortium for Lexical Research:

Ftp File: pmtex-1.2.tar.Z
Short Description: Simple (La)TeX system for typesetting Chinese, Japanese, and Korean

Ftp Directory: pub/typesetting/tex
More Info: info/11

-+

Copyright notice: pmtex

-+

[CLR Note: this description was extracted from the top-level "README" file in the pmtex archive file.]

Poor Man's TeX

This is the usual in-the-absence-of-real-documentation readme file for Poor Man's Chinese and Japanese.

pmC and pmJ are less than ideal implementations of Chinese and Japanese for TeX. Less than ideal because they use fonts based on 24x24 dot-matrix fonts, and don't do vertical format typesetting and so forth. However, they (seem to) work, are free, and work with a standard TeX of version 3 with no known system dependencies.

LEGALITIES

Portions of pmC and pmJ are copyrighted free software. See the file license for details. The TeX portions are public domain.

Original Author: ridgeway@blackbox.hacc.washington.edu (Thomas Ridgeway)
Modifications/Additions: mleisher@nmsu.edu (Mark Leisher)

EVALUATION OF NYSERNET'S NEW CONNECTIONS PROGRAM

Submitted to
ASIS Midyear Meeting 1992

Ann P. Bishop
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801
bishop@alexia.lis.uiuc.edu
(217) 244-3299

Philip Doty
Graduate School of Library and Information Science
University of Texas at Austin
Austin, TX 78712-1276
pdoty@utxvm.cc.utexas.edu
(512) 471-3839

Charles R. McClure
School of Information Studies
Syracuse University
Syracuse, NY 13244-4100
cmcclure@suvm.acs.syr.edu
(315) 443-2911

Abstract

NYSERNet, a mid-level regional network, has instituted the New Connections Program, providing free dial-up connections to the NYSERNet backbone for particular kinds of institutions in New York State. These connections are available to communities with limited experience in networking: museums, K-12 schools, government agencies, libraries, library consortia, and small colleges. NYSERNet initiated the pilot program in order to challenge and assist new communities of users to develop programs and services that would take advantage of public high-speed networks. NYSERNet also used the program to gain a clearer understanding of its clients and potential clients' needs and to prepare for the planning, implementation, and evaluation of future user-based programs.

The New Connections Program is important as a test of innovative roles for mid-level networks. It is also important as an experiment in the development and evaluation of networking applications by new user communities themselves.

This paper will briefly report the results of the preliminary evaluation of the New Connections Program that included:

- The identification of institutions' objectives for their participation in the New Connections Program, how those objectives related to the institutions' missions, and the development of appropriate criteria for the evaluation of the Program

- Description of the networking activities of the experiment's participants -- e.g., who used the connection? For what purposes?

The paper will also describe the recommended full evaluation of the Program that will include surveys of and interviews with New Connections Program participants and logs of the activities of both participating institutions and support institutions.

The paper will emphasize the process of developing appropriate methodologies for the evaluation of this type of network service and will discuss the role of user-based evaluation in helping to provide more responsive and successful networking services.

Introduction

NYSERNet's New Connections Program is an experiment funded in part by the New York State Science and Technology Foundation and the National Science Foundation. It offers dial-up modes of connection to the NYSERNet backbone for trial periods and, through that backbone, to the Internet. The network services provider is Performance Systems International (PSI) of Reston, Virginia. The aim of the program is to offer network connections to new user communities, e.g., museums, government agencies, K-12 schools, libraries, library systems, and small colleges, so that they may take advantage of the full range of Internet services. Participating institutions are provided with the trial membership in NYSERNet at no cost as well as user support documentation and training services, including those offered by PSI.

In turn, participants in the program provide appropriate hardware (computer and modem) and a standard phone line. Each institution designates an individual to serve as the New Connections liaison, whose role is to set up the equipment and mediate between users and service providers. Institutional liaisons include resident computer experts or simply those individuals at participating institutions who are interested in the Program and agree to take an active role. Users of New Connections services also agree to report on the use and utility of their network connections.

This paper outlines a plan for the evaluation of the New Connections Program. Some preliminary evaluation has been done and has resulted in a number of products to support the full program evaluation. These products are described below. The full evaluation will include the collection, analysis, and reporting of pertinent network use and evaluation information from New Connections participants. It will also include the collection and analysis of information obtained from NYSERNet and PSI. The final evaluation report will offer recommendations to NYSERNet about the New Connections Program and related user-based initiatives.

Goals of Evaluation

High-performance computing and networking services are of great importance to the United States (Office of Science and Technology Policy, 1990; U.S. Senate, 1990). In order to maximize the value and utility of networking services, they must be designed with the needs and goals of their users in mind (McClure, Bishop, Doty, & Rosenbaum, 1991; Panel on Information Technology and the Conduct of Research, 1989). Given this, the user-based evaluation of networking services is of critical importance.

The process of evaluation described here can contribute to the development of improved networking services by examining in detail the outcomes of one innovative networking program. At present, little is known about network use, beyond the most basic counting of traffic and machines. This ignorance is especially severe among the communities that the New Connections Program aims to serve.

A user-based evaluation identifies three essential characteristics of a community and its use of an information resource or service:

- What is the specific audience for the resource or service?

- What are the goals, skills, and social setting and structure of that audience?

- How does the resource or service help achieve those goals, support and extend those skills, and fit into that social setting?

Few user-based evaluations of network services have been done, especially for the communities served by the New Connections Program. User-based evaluations of network services and products must be systematic, empirical investigations of network users' behavior, expectations, and success and failure factors. Often, users find it difficult to integrate advanced information technologies into their work. Many evaluations place an overemphasis on system features at the expense of addressing issues related to users' attempts to integrate the new technology into their daily lives. They describe what is used, but not why, how, or to what end. The evaluation of the New Connections Program will try to investigate both technical and social factors related to system use and will describe networking goals, problems, and outcomes as experienced by network users.

The evaluation also makes a conscious attempt to consider the perspectives and concerns of both the Program participants and the providers of Program services. Evaluation of the users' networking activities and their perceptions of the value of these activities will focus on such questions as:

- Did the New Connections Program allow participating institutions and individual users to accomplish their stated objectives for the network connection?

- What were the impacts of the connection on the participating organizations and individuals in those organizations?

- What factors, e.g., program training and support, organizational culture and resources of participating institutions, and the nature of participants' goals, contributed to program success?

- What were the major obstacles to the success of the New Connections Program *from the point of view of the users*? Were these technical, financial, personal, legal, etc.? How might such problems be alleviated in the future?

The evaluation of the Program from NYSERNet's perspective will emphasize similar questions, but would also include others:

- Did the New Connections Program accomplish NYSERNet's goals for this initiative? How did the Program fit into NYSERNet's overall programmatic strategy?

- What were the major obstacles to the success of the New Connections Program *from the point of view of NYSERNet*? Were these technical, financial, personal, legal, etc.? How might such problems be alleviated in the future?

- Were the major participants (individual institutions, NYSERNet, and PSI) pleased with their relationships and responsibilities, especially for training, documentation, and user support?

- What additional kinds of institutions (e.g., other non-profit organizations and government agencies) should be represented in this kind of intitiative?

Another major goal of the assessment will be the provision of specific and general recommendations to NYSERNet for the improvement of its networking programs and services. The evaluation process will also serve as a laboratory for determining appropriate strategies, instruments, and activities for such a user-based network evaluation, and it will serve as a model to be improved and adapted to different programs, institutions, and circumstances.

Preliminary Evaluation Activities

Preliminary work for the evaluation of the New Connections Program was done during the summer and fall of 1991. This preliminary work achieved a number of results. Data collection instruments for the full evaluation were developed, including:

- Forms for participants' preliminary statements about their goals and expectations for their network connections (see Appendix A). These questionnaires were sent to all Program liaisons.

- Forms for particpants' final statements about their use of the connection, its utility, and how well the program matched their expectations (see Appendix B)

- Guides to be used in interviews with program participants (see Table 1)
- User logs (see Table 2)
- Liaison diaries
- Guides for service providers to record their interactions with participants.

These instruments are only early iterations; they will be pretested and changed as appropriate. In addition to the development of these instruments, the preliminary stages of teh evaluation included the partial analysis of the statements of the program participants, as described below.

There are twenty-nine institutions participating in the New Connections Program, and twenty-three returned preliminary statements to the study team early enough to be included in this summary. The following summary is based on those twenty-three documents.

The institutions participating in the New Connections Program fall into four major groups, with a number of sub-categories (see Table 3). In the numbers cited below, if only one number is given, that number represents the number of participants in that category, all of whom returned their preliminary statements. If two numbers are given, the last represents the number of participants in that category, while the first represents the number of participants who returned their preliminary statements.

In response to several survey questions, liaisons gave a number of reasons given for participating in the New Connections Program (Table 4). Almost every respondent specifically noted the desire to increase the awareness of the existence and benefits of networked information services in their institution. This desire was coupled with the aim of increasing institutional use and knowledge of online services, especially for those members of the institution sceptical of the benefits of networking.

Note that goals were sometimes expressed simply in terms of networking functions, such as email, that participants wanted to use; other goals were stated in terms of outcomes related to the nature of work performed at the institution.

In the educational institutions, particular benefits, especially the provision of information for research projects and expanded communication possibilities, were expected for both students and teachers. One Program liaison, while discussing curriculum development, said that the NYSERNet connection was expected to help only mathematics and science teachers. This response demonstrates a common misconception and prejudice about the use of computing and networks -- the belief that they are intended for and valuable to only those involved in mathematics and the physical sciences.

The preliminary survey also asked participants to identify their criteria for the success of the New Connections Program (Table 5). Responses were usually more institution-specific than the benefits expected noted above. Many of the criteria noted were, at the same time, very general, with no apparent way to operationalize or specify the criteria mentioned. There were two respondents who did not answer the question about criteria for success, so the summary below is based on N=21.

In addition to (or instead of) stating evaluation criteria, some respondents noted the mechanisms by which the information needed to assess the degree to which their criteria were met would be gathered. These included questionnaires, interviews, consultations, user logs, and analysis of the content of network communications.

This last mechanism might be a bit disturbing for its implications for user privacy. At the same time, however, it may refer only to examination of documents clearly intended to be public. This response does indicate that some socialization into the role of network/information intermediary and explicit discussion of appropriate and inappropriate gatekeeping behaviors may be an important part of what users and liaisons need from NYSERNet. At the same time, however, such socialization may be difficult to achieve in certain organizations.

The preliminary phase of the program assessment helped to focus the study team's attention on how specifically to elicit and understand participants' goals and expectations, how the participating institutions and users would judge the success or failure of the network connections, and what data collection and analysis techniques would be appropriate and useful in the full evaluation.

The evaluation process described here is <u>user-based</u> in that it evaluates the effectiveness of the networking service provided from the point of view of its users. In other words, particular attention is given to the degree to which the connection and services offered matched users' needs, goals, expectations, and settings (Dervin & Nilan, 1986; Galegher, Kraut, & Egido, 1990; Hiltz, 1984; Sproull & Kiesler, 1991; Taylor, 1991). This kind of user-based focus is important for several reasons. The user communities represented in this program (e.g., K-12 schools, museums, small colleges, libraries, and library systems) have only recently been of interest to network service providers; thus, these kinds of institutions' particular concerns, problems, and institutional missions have rarely been of interest to investigators of networking. For example, the use of electronic networks to improve education, as opposed to simply "automating" existing educational practices, has been noted by educational researchers (e.g., Kay, 1991; Riel, 1986), but deserves greater attention. In addition, we know relatively little about the users and uses of network services in terms of motivation, problems, attitudes, and expectations. Finally, there is no established method for bringing user-based evaluations to the development of network policies and services. Therefore, there is little methodological guidance available about which questions are important in such an evaluation study, how to analyze the data collected, and how best to use the study's results in order to develop network services that support users' tasks and goals.

Plan for Program Evaluation

The major activities involved in the full evaluation of NYSERNet's New Connections Program will include:

- Gathering and analyzing weekly diaries maintaned by New Connections liaisons and user logs (Table 2) that track network activities and general experiences in the New Connections Program.

- Gathering and analyzing participant reports of, for example, network activities and uses, the type and number of users, perceived impacts, problems, equipment used, and specific and general recommendations for the Program (see Appendix B).

- Undertaking site visits, interviews, and/or focus group discussions with selected participants to collect additional data about users' networking activities, expectations, and attitudes. The interviews and focus groups will take place at the participants' home institutions or at NYSERNet facilities. Part of the aim of these data collection activities will be to visit at least one of each of the kinds of institutions represented in the Program, e.g., museums, primary schools, secondary schools, small colleges, libraries, etc. (Table 1).

- Gathering and analyzing PSI logs of their telephone conversations with program users

- Interviewing, by telephone, PSI support staff involved in the Program

- Interviewing NYSERNet representatives to determine their goals for the New Connections Program and how that program relates to NYSERNet's general mission and other programmatic initiatives e.g., the Bridging the Gap program, the Academic Scholar Access Program (ASAP), and the already-functioning K-12 and library interest groups and their initiatives (BOCES/RIT Project, C.R.E.S.T. Project, and the SNAP-NY Project).

- Collection and analysis of data related to NYSERNet's direct support of users in the Program by interviewing the NYSERNet staff involved, examining the appropriate transaction log(s), and correlating these sources with data from the users and from PSI.

- Encouraging NYSERNet and PSI to set up an electronic bulletin board system for New Connections particpants, PSI, and NYSERNet. The purposes of the bulletin board is to allow New Connections institutions to share their experiences, frustrations, and successes; encourage them to use the network to develop their networking skills and solve their own problems; help the

participants to develop realistic expectations about their network connections and networking in general; give NYSERNet and PSI insight into specific user problems; convince participants that the service providers care about the success of the program; and provide the study team with additional data about users' expectations, network uses, and problems.

- Producing a written report to NYSERNet to include:

  -- A description of participants' network activities

  -- An overall assessment of the New Connections Program

  -- Recommendations about how such programs can be made more effective.

Conclusions: Outcomes and Benefits of the Evaluation

The evaluation of the NYSERNet's New Connections Program will produce significant benefits for NYSERNet and other members of the networking community. It will:

- Provide a record of the activities and experiences of New Connections participants
- Allow a better understanding of networking clients' needs, activities, perceptions, and behavior
- Offer new insights into goals and operations of networking services and programs
- Better prepare service providers to initiate, manage, and evaluate future user programs
- Produce evaluation instruments to serve as models for other networking evaluation efforts
- Give service providers experience in the conduct of user-based research.

An evaluation such as that planned for the NYSERNet New Connections Program gives the network service provider the opportunity to raise its profile in the provision of user support and innovative programs. Moreover, the proposed evaluation will offer NYSERNet and other networking service providers a range of possible strategies for developing new services and operations and increasing their overall effectiveness.

The evaluation will also be a valuable exploration of how mid-level networks can support users and act as effective liaisons between them and the Internet. This area will assume greater importance as networks evolve and as the size and heterogeneity of the networking community increases.

One of the most important outcomes of the proposed evaluation is that NYSERNet will gain experience with the conduct of user-based research in the study of electronic networking. The proposed study will serve as a laboratory for the identification, development, and refinement of approppriate methodologies for doing this sort of evaluation and using the evaluation to inform networking services and policy decisions.

Bibliography

Dervin, Brenda, & Nilan, Michael. (1986). Information needs and uses. In Martha Williams (Ed.), Annual review of information science and technology (Vol. 21, pp. 3-33). Chicago, IL: Knowledge Industry Publications.

Galegher, Jolene, Kraut, Robert E., & Egido, Carmen. (Eds.). (1990). Intellectual teamwork: Social and technological foundations of cooperative work. Hillsdale, NJ: Lawrence Erblaum Associates.

Hiltz, Starr Roxanne. (1984). Online communities: A case study of the office of the future. Norwood, NJ: Ablex.

Kay, Alan C. (1991). Computers, networks, and education. Scientific American, 265(3), 138-148.

McClure, Charles R., Bishop, Ann P., Doty, Philip, & Rosenbaum, Howard. (1991). The National Research and Education Network (NREN): Research and policy perspectives. Norwood, NJ: Ablex.

Office of Science and Technology Policy. Federal Coordinating Council for Science, Engineering, and Technology. Committee on Physical, Mathematical, and Engineering Sciences. (1991). Grand challenges: High Performance Computing and Communications, the FY 1992 U.S. research and development program. Washington, DC: Office of Science and Technology Policy.

Panel on Information Technology and the Conduct of Research. (1989). Information technology and the conduct of research: The user's view. Washington, DC: National Academy Press.

Riel, Margaret. (1986). The educational potential of computer networking. Interactive Technology Laboratory Report #16. ED 311877.

Sproull, Lee, & Kiesler, Sara. (1991). Connections: New ways of working in the networked organization. Cambridge, MA: MIT Press.

Taylor, Robert. (1991). Information use environments. In Brenda Dervin & Melvin J. Voigt (Eds.), Progress in communication sciences (Vol. 10, pp. 217-255). Norwood, NJ: Ablex.

U.S. Senate. (1990). High-Performance Computing Act of 1990. Committee on Commerce, Science, and Transportation. Report 101-387. Washington, DC.

Table 1. Interview Guide

Current status of your New Connections link

1. Tell us about how you (the liaison and others at the institution) are using your NYSERNet connection.

2. What specific types of network services have you used or tried to use? What sorts of information are you interested in?

3. What kinds of problems and issues have you encountered?

4. What benefits are you experiencing as a result of the New Connections program?

Your assessment of the New Connections Program

1. Is the Program working out the way that you expected?

2. Tell us about being a liaison. What are the benefits and problems you experience as an individual?

3. How do you feel about the support you've received for participating in the New Connections Program (informational, technical, moral, etc.)?

4. What's your general assessment of the Program so far?

5. What suggestions would you offer for improving the New Connections Program?

6. Do you think that a program like this is a good way to find out about the needs and interests of user groups new to networking? Is it a good way to encourage users to become regular NYSERNet clients?

Table 2. User Log of New Connections Activities

Institution Name: ____________________

Date	Your Institutional Role	Network Function Attempted	Your Response

Table 3. Participating Institutions

Educational Institutions (14/18)	
Primary	4
Primary and Secondary	1
Intermediate	1
Secondary	3
School Districts	1
Four-year Colleges	4/8
Libraries (7/9)	
Public	2
Library Systems	3/5
Hospital	1
State Library	1
Government Agencies (1)	1
Other non-profit Institutions (1)	1
TOTAL	23/29

Table 4. Reasons Cited for Participation

Increased awareness of network benefits	21
Email, especially access to bulletin boards and listservs	16
Information for research projects	13
Searching other databases (including Federal and state databases)	10
Searching online public access catalogues (OPACs)	7
Curriculum development	7
File transfer	4
Document delivery, resource sharing, and interlibrary loan	3
Development of networked instruction, especially for distance education	3
Student publishing	1
Access to specific software	1
Establish elecronic mentor program between research scientists and primary students	1
Make teachers use the information center/library more	1
Reliable network connection	1

Table 5. Respondents' Criteria for Program Success

Usage level, including number of users and time spent online	11
Positive user reactions	6
Resources to be reached and success in reaching them	3
Getting staff and patrons "interested in global possibilities"	2
Reliability of connection	2
Quality of "students' finished products"	2
Ease of installation	1
Integration into curriculum development	1
Ease of logon and use of OPACs	1
Ease with which children can get online and communicate with others	1
Ease of use of Internet vs. "other ways to get the same information"	1
Cost of connecting to and retrieving information	1
"Was it fun and exciting?"	1

NYSERNet New Connections Program
Preliminary Participant Statement

We would like to know a little bit about your organization and your plans for participating in the New Connections program. The information you provide will help us to understand your needs and evaluate the strengths and weaknesses of the program. Feel free to attach extra sheets if the space provided for any answer is not sufficient.

A LITTLE BIT ABOUT YOU AND YOUR INSTITUTION

1) Your name: ______________________ Phone: ____________
2) What is your current job title or position? ______________________
3) Name of institution: ______________________

4) Type of institution:

__ Primary education
__ Secondary education
__ 2-Yr. College
__ 4-Yr. College or University
__ Post-graduate education
__ Museum
__ Library
__ Library system
__ Other: ______________________

5) How experienced are you as a network user?

__ Not at all __ A little __ Somewhat __ Very

6) What are some of the electronic networks and applications (e.g., electronic mail on BITNET) people at your institution currently use?

YOUR CONNECTION STATUS

7) Connection status as of ________ (fill in today's date):

__ We have not yet started the connection process.
__ We have started the connection process, but we are not yet operational.
__ Our connection is complete and operational, but we're not using it yet.
__ We've been using our connection since: __________ (fill in approximate date of first use).

8) If you are already connected:

a) Which type of access do you have?

__ PCMAIL __ POP __ UUPSI __ Don't know

b) Are you connected to USENET/NEWS?

__ Yes __ No __ Don't know

YOUR GOALS AND EXPECTATIONS

Please help us understand your expectations and motivations regrading the New Connections program.

9) By the end of the program, who do you hope will be using your network connection?

__ Teachers
__ Students
__ Librarians
__ Administrators
__ Support staff
__ Others: ____

10) What activities do you think people at your institution will use networks to support (e.g., access to outside people who share their interests, curriculum development, the dissemination of institutional information)? Please be as specific and complete as possible.

11) By the end of the program, about how many people at your institution do you hope will be using your connection? ____

12) How will potential users at your institution find out about the New Connections program?

13) Which networking applications do you think people at your institution will be most anxious to use?

__ Electronic mail
__ File transfer
__ Electronic bulletin boards, mailing lists, etc.
__ Information retrieval, including online public access catalogues (OPACs)
__ Access to data sets
__ Access to remote computers (e.g., supercomputers)
__ Other: ____

14) How will they learn to use these applications?

15) How would you describe the current level of networking and computing expertise of intended New Connections user group(s)?

16) Why did you join the New Connections program? In general, what benefits do you hope will result from your institution's participation?

17) What criteria will you use to decide whether your New Connection experience was successful or not?

18) What do you expect that acting as the New Connections liaison will entail for you individually?

USER SUPPORT

19) What problems do you think you and others at your institution might experience as New Connections participants?

20) What problems or barriers have you experienced already?

21) What kind of expertise is available at your institution to help set up and use the network?

22) What kind of assistance do you think you'll need to use your New Connection successfully?

__ Technical support related to hardware
__ Instructions for using the network, e.g., logging on, sending mail, transferring files
__ Directories and guides to people and resources available online
__ General information about what networks can be used for
__ Other (please specify):

23) Where do you expect to get this assistance? *Please circle the kind(s) of assistance you expect from each of the following sources.*

Own institution
Tech Support / Instructions / Directories / General Info / Other: ____________

Personal contacts outside your institution
Tech Support / Instructions / Directories / General Info / Other: ____________

Other New Connections participants
Tech Support / Instructions / Directories / General Info / Other: ____________

PSI
Tech Support / Instructions / Directories / General Info / Other: ____________

NYSERNet
Tech Support / Instructions / Directories / General Info / Other: ____________

Other sources of assistance: ______________________
Tech Support / Instructions / Directories / General Info / Other: ____________

24) Do you have any other comments you would like to make about the New Connections program?

Thank you for your help. If you have any questions about this survey or if you would like to offer further comments, please call Ann Bishop or Philip Doty at (315) 443-2911.

Please use the enclosed pre-addressed envelope to return your completed questionnaire BY AUGUST 30th.

NYSERNet New Connections Program
Participant Final Statement

We would like to know more about your organization and your participation in the New Connections program. The information you provide will help us to understand your needs and evaluate the strengths and weaknesses of the program. Feel free to attach extra sheets if the space provided for any answer is not sufficient.

YOU AND YOUR INSTITUTION

1) Your name: ______________________________ Phone: ____________
2) What is your current job title or position? ______________________________
3) Name of institution: ______________________________

4) Type of institution:

__ Primary education
__ Secondary education
__ 2-Yr. College
__ 4-Yr. College or University
__ Post-graduate education
__ Museum
__ Library
__ Library system
__ Other: ______________________

5) How experienced are you as a network user?

__ Not at all __ A little __ Somewhat __ Very

YOUR CONNECTION STATUS

6) We've been using our connection since: __________ (fill in approximate date of first use).

7) a) Which type of access do/did you have?

__ PCMAIL __ POP __ UUPSI __ Don't know

b) Are/were you connected to USENET/NEWS?

__ Yes __ No __ Don't know

USES

Please help us understand the uses made of your NYSERNet connection.

8) By the end of the program, who was using your network connection?

__ Teachers
__ Librarians
__ Support staff
__ Others: ______________________________
__ Students
__ Administrators

9) What activities did people at your institution use networks to support (e.g., access to outside people who share their interests, curriculum development, the dissemination of institutional information)? Please be as specific and complete as possible.

10) By the end of the program, about how many people at your institution were using your connection?

----

11) How did potential users at your institution find out about the New Connections program?

12) Which networking applications did people at your institution use?

__ Electronic mail
__ File transfer
__ Electronic bulletin boards, mailing lists, etc.
__ Information retrieval, including online public access catalogues (OPACs)
__ Access to data sets
__ Access to remote computers (e.g., supercomputers)
__ Other: ____

13) How would you characterize the level of use of the network connection in your institution?

__ Extensive __ Moderate __ Light __ None

14) What were the most valuable features of the connection?

15) How did users learn how to use your network connection and applications?

16) What benefits resulted from your institution's participation in the New Connections Program?

17) Did your organization experience any negative impacts from the Program? If yes, what were they?

18) How successful was your New Connection experience according to the following criteria?

Effort required from you

|____________|__________|____________|____________|

Extremely successful *Not at all successful*

Cost (in $)

|____________|__________|____________|____________|

Extremely successful *Not at all successful*

Time required

|____________|__________|____________|____________|

Extremely successful *Not at all successful*

Number of users

|____________|__________|____________|____________|

Extremely successful *Not at all successful*

Types of users

|____________|__________|____________|____________|

Extremely successful *Not at all successful*

Other criteria (Please specify)

____

|____________|__________|____________|____________|

Extremely successful *Not at all successful*

____

|____________|__________|____________|____________|

Extremely successful *Not at all successful*

19) Were users at your institution, including you, able to do things that you did not expect to be able to do with your network connection?

a) What were they?

b) Why was it a surprise?

20) Were users at your institution, including you, unable to do things that you did expect to be able to do with your network connection?

a) What were they?

b) Why weren't you able to do them?

21) What did acting as the New Connections liaison demand from you individually?

22) What equipment did you use for the New Connections Program? Please include all computers, modems, etc.

USER SUPPORT

23) What problems did you and others at your institution experience as New Connections participants?

24) What kind of assistance did you need in order to use your New Connection successfully?

__ Technical support related to hardware
__ Instructions for using the network, e.g., logging on, sending mail, transferring files
__ Directories and guides to people and resources available online
__ General information about what networks can be used for
__ Other (please specify):

25) Did you get this assistance?

26) If you got this assistance, where did you get it? *Please circle the kind(s) of assistance you got from each of the following sources.*

Own institution
Tech Support / Instructions / Directories / General Info / Other: ____________

Personal contacts outside your institution
Tech Support / Instructions / Directories / General Info / Other: ____________

Other New Connections participants
Tech Support / Instructions / Directories / General Info / Other: ____________

PSI
Tech Support / Instructions / Directories / General Info / Other: ____________

NYSERNet
Tech Support / Instructions / Directories / General Info / Other: ____________

Other sources of assistance: ______________________
Tech Support / Instructions / Directories / General Info / Other: ____________

26) What were the major strengths of the New Connections Program?

27) What were the major weaknesses of the Program?

28) Do you have any other comments you would like to make about the New Connections program?

Thank you for your help. If you have any questions about this survey or if you would like to offer further comments, please call XXXXXX at (315) 443-2911.

Please use the enclosed pre-addressed envelope to return your completed questionnaire BY XXXXXX.

CULTIVATING THE ELECTRONIC HEARTLAND: PREPARING FOR THE COMING KNOWLEDGE HARVEST

by Arthur J. Murray

ABSTRACT

The interconnection of information resources via national and global computer networks has resulted in the creation of a worldwide distributed knowledge base. The capability to instantly access this vast information resource presents unlimited opportunities for the acquisition and dissemination of knowledge. Just as industry has used automation to relieve much of its manual labor burdens and increase productivity, reliable tools are needed in order to exploit the many varieties of available networked resources. This paper describes the results of the author's research in developing tools to support: 1) knowledge acquisition (for the network user in search of knowledge), and 2) knowledge dissemination (for the user wanting to share knowledge).

Functional requirements have been generated based on observations of difficulties encountered by users of internetworked resources. A prototype electronic knowledge harvester was developed by integrating artificial intelligence, data base, and communications technologies. Tests were performed to demonstrate the effectiveness of knowledge-based network interfaces in supporting wide area network information search, access and retrieval. Preliminary results show that speed-up improvements of at least one hundred percent are possible in locating and accessing internetworked resources. Techniques for the dissemination of knowledge show similar preliminary results. Remaining issues for continued investigation are identified.

WHAT THE NEW ELECTRONIC HEARTLAND HAS TO OFFER

National and global internetworking initiatives have in essence created a worldwide distributed knowledge base. This knowledge base is vast, interconnecting innumerable host systems and information services. Some of the many types of information services currently available include news wire feeds, on-line data bases, electronic bulletin boards and electronic shopping malls, as well as processing services on host systems to support remote program execution. The proliferation and interconnection of such a wide variety of knowledge sources and services is responsible for the creation of what is being referred to as the new electronic heartland (Naisbitt, 1990).

The benefits of tapping such a vast array of resources are many, and new knowledge and applications are continually unfolding. Knowledge can be acquired and exchanged from the home, office, hotel, airplane, or automobile. Collaborative research and problem solving activities are taking place across international boundaries on a regular basis. In addition, competition among carriers and information services providers continues to drive

telecommunications, networking and processing charges downward. As a result, one-stop global knowledge harvesting has become a practical and attractive option for many users of information systems.

With this rich, mostly untapped electronic resource, the potential exists for the near instantaneous harvesting of new varieties of knowledge crops. These crops could range from varieties that grow wild to genetically engineered hybrids. As an example of harvesting knowledge of the wild variety, a user could set broad search criteria that would look for little known relationships or discrepancies among a finite set of knowledge elements. For example, a systematic search might uncover investment opportunities created by favorable spreads in theoretical values between two different financial instruments, such as the market value of a foreign currency and the price of gold in that currency. Such electronically harvested knowledge would give an investor an edge that might not otherwise be available.

An electronic knowledge harvester could be used to extract data from several on-line data bases, load the data into a statistical package for analysis, and output the results. For problems of a broader scope, genetically engineered knowledge could be produced by using genetic algorithms to identify and evaluate candidate solutions (Goldberg, 1989).

The most exciting potential of the electronic heartland lies in the creation of social interactions cultivated by synergistic use of the internetworking facility. By automatically collecting and storing knowledge regarding network users and their areas of expertise, researchers from around the world having similar interests can be brought together. For example, a researcher in Eastern Europe, interested in determining impacts of transitioning fuel production from government to private enterprise, may be able to identify a knowledgeable source from a country where a similar transition has already taken place.

Presently, the accomplishment of these tasks requires a great deal of effort, and there are no guarantees of success. If anything, uncovering specialized knowledge or identifying individuals with unique interests usually happens purely by accident---a stock market analyst stumbles upon an undiscovered pattern in a series of price charts, or a researcher happens to meet somebody on the subway that provides a lead into identifying a potential collaborator for a project.

In order to fully exploit the capabilities of internetworking, the process of knowledge harvesting must be systematic, not accidental. By incorporating more intelligence into network interfaces, the system can relieve the human user of many of the burdens of trying to keep track of where everything is on the network. This will leave the user free to explore questions and possibilities, while the system interprets those needs and helps open the doors that lead to new discoveries.

REQUIREMENTS FOR CULTIVATING KNOWLEDGE ON A LARGE SCALE

The goal of the electronic knowledge harvester is to establish a means whereby knowledge seekers and knowledge providers can be put in touch with each other through an electronically mediated process. Because international boundaries are involved, the process must take into account appropriate tariffs and government regulations.

There are several ways this goal can be achieved. In one case, knowledge seekers need to locate the sources of the knowledge they are seeking. Knowledge sources are data bases or information services that maintain knowledge, but do not actively look for specific knowledge seekers. The knowledge seekers must come to them.

Knowledge providers, on the other hand, have knowledge they want to share, but do not always know with whom they should share that knowledge. There may exist users that could benefit from specific pieces of new knowledge, but they are not actively seeking that knowledge. Knowledge providers must come to them.

In order to better understand requirements for the electronic exchange of knowledge, knowledge elicitation sessions were conducted with users of internetworked information systems. The data collection process consisted of both interviews and the electronic logging of user-machine dialogs generated during typical internetwork sessions. Several obstacles preventing the full use of the knowledge available within the electronic heartland were identified in the process. One major shortfall is that the knowledge seekers do not always know what knowledge is available or where to find it. Conversely, the knowledge providers do not always know how to locate or get the attention of those in need of the knowledge they have already harvested.

Even if the knowledge seeker and knowledge provider know of each other, there are often too many demanding manual burdens that inhibit the process whereby knowledge can be exchanged. These burdens are briefly described in the following paragraphs.

<u>Limited on-line help facilities</u>. Most of the on-line help facilities were found to be descriptive rather than prescriptive in nature. They tell the user about the system but give little direction regarding how to proceed.

<u>Lack of a one-stop facility for identifying available network resources</u>. Users spent a majority of their time fumbling through post-it notes, loose-leaf binders, newsletters and other fragmented documentation of candidate knowledge sources.

<u>Excessive amounts of guesswork required to establish a connection and conduct a meaningful dialog</u>. Users applied a great deal of trial and error in order to figure out the correct communication protocols and command languages needed to interact

with a host system.

Excessive time wasted on trial and error searching. This problem results in wasted expenditures incurred from users conducting searches that ultimately lead to dead-ends.

Being in the right church but the wrong pew. Users often encountered the added frustration of finding out that the correct knowledge had been available through one or more of the sources that were accessed, but the user had given up the search too quickly.

Other obstacles inhibiting getting knowledge crops to market include:

1) lack of protocol standardization
2) issues concerning intellectual property versus knowledge dissemination
3) determining restrictions, surcharges, and appropriate licensing and royalty requirements.

Based on an examination of these shortfalls, a formal set of functional requirements was developed. The requirements are broken down into the following three categories:

1) knowledge acquisition requirements
2) knowledge dissemination requirements
3) knowledge base maintenance requirements.

The requirements are summarized in Figure 1 and are described in the three subsections that follow.

Knowledge Acquisition Requirements. A clear and accurate description of the problem under investigation by the knowledge seeker is of foremost importance in order to achieve successful knowledge acquisition. The problem statement must be decomposed and matched against the attributes of available knowledge sources. An assessment is then required in order to identify and prioritize the most promising candidates for searching.

In the next step, the candidate knowledge sources must be accessed, queries must be generated, and the knowledge, if available, must be retrieved. This must take place in as expeditious a manner as possible in order to minimize connect charges. However, the process is iterative and several passes through identical knowledge sources may be required. In addition, the new knowledge that is retrieved may cause the user to rethink the problem, thereby causing a return to the problem definition portion of the process.

Knowledge Dissemination Requirements. The requirements for knowledge dissemination are very similar to those for knowledge acquisition. A clear and accurate description of the scope, activities, problem and solution for each candidate for dissemination must be formulated as a highly structured case. Each case must then be decomposed into problem-solution pairs. The attributes of each pair must be matched against attributes

KNOWLEDGE ACQUISITION REQUIREMENTS:

- PROBLEM DEFINITION
- IDENTIFICATION & PRIORITIZATION OF CANDIDATE KNOWLEDGE SOURCES
- ACCESS TO KNOWLEDGE SOURCES
- QUERY GENERATION
- KNOWLEDGE RETRIEVAL
- LOGOFF

KNOWLEDGE DISSEMINATION REQUIREMENTS:

- CASE DEFINITION
- IDENTIFICATION & PRIORITIZATION OF CANDIDATE KNOWLEDGE RECIPIENTS
- ACCESS TO KNOWLEDGE RECIPIENTS
- REPORT GENERATION
- KNOWLEDGE DISSEMINATION
- LOGOFF

KNOWLEDGE BASE MAINTENANCE REQUIREMENTS:

- ESTABLISH AND MAINTAIN DEFAULT SEARCH CONSTRAINTS
- SPECIFY PERMITTED DEGREE OF SYSTEM AUTONOMY
- MAINTAIN YELLOW PAGES
- MAINTAIN PROTOCOL AND COMMAND FILE LIBRARIES
- MONITOR USAGE RATES AND CHARGES

Figure 1. Requirements for an Electronic Knowledge Harvester

describing the areas of interest of potential knowledge recipients. An assessment is then required in order to identify and prioritize the most promising candidate recipients. The candidate recipients must then be accessed, reports generated, and the knowledge disseminated as expeditiously as possible. Feedback from recipients concerning the usefulness of the knowledge can be used to further refine the dissemination process.

Knowledge Base Maintenance Requirements

This set of requirements deals with maintaining the knowledge bases for both seekers and providers. The first requirement is to establish and maintain a set of default constraints to be used in controlling the search process. These constraints include:

1) prioritized subject areas and topics of interest
2) timeliness and perishability factors
3) cost limitations.

The candidates for search are maintained in an intelligent yellow pages directory. This portion of the knowledge base contains the topics covered by known knowledge sources and potential recipients, along with access charges, including discounts for usage during off-peak hours. The knowledge base must also contain the telecommunications links by which each known source or potential recipient can be accessed. For each link, the appropriate electronic addresses, access procedures, communication protocols and system commands must be maintained.

The knowledge base must be able to perform inference operations in order to arrive at an approach to fulfilling a request either for retrieval or dissemination. For instance, timeliness and perishability constraints must be weighed against access charges (i.e., whether the candidate knowledge sources must be searched immediately or if the process can be postponed until a time period when usage rates are more favorable).

The amount of autonomy the knowledge harvester will have also needs to be determined. This defines the latitude the system will be given with regard to the extent of the search to be conducted, and the minimum acceptable degree of similarity between the attributes of seekers and known sources, or between providers and potential recipients.

Finally, the knowledge base will be required to maintain and analyze an automatic log that can be used to evaluate, over a given period of time, the procedures having the highest rate of success. Success occurs when the needed knowledge is retrieved or disseminated within the specified time frame at the lowest possible cost.

Many of these requirements can be satisfied by the skillful and innovative application of information technology. Successful technology application will result in significant improvements over current, manually intensive processes. These include

improvements in:

1) consistency
2) speed
3) cost effectiveness
4) locating the correct knowledge sources/recipients
5) enforcement of appropriate royalties, licensing fees and dissemination restrictions.

The next section discusses ways that today's technology can be applied in order to achieve these results.

THE KNOWLEDGE HARVEST TOOLSHED

The electronic knowledge harvester is an intelligent internetwork interface that mediates the knowledge acquisition and dissemination processes. Many different interfaces are possible, ranging from manual to totally autonomous. Most of today's information systems have manual interfaces, in which the user must perform a majority of the steps needed to access and conduct dialogs with services on the network. At the other end of the spectrum are totally autonomous interfaces, in which the system has enough intelligence to interpret the user's requests, access and interact with internetworked knowledge sources, interpret the network responses, and present the user with the final results. The knowledge harvester described in this paper is semi-autonomous, in which the machine interprets and acts upon user requests, but the user must interpret and act upon network responses.

Figure 2 provides a block diagram of the tools that were used to build a prototype semi-autonomous interface. The user enters a request in english. A natural language interpreter converts the request into structured query language (SQL) statements. The SQL statements are then evaluated by a knowledge-based system in two ways. First, a search is conducted by the local data base to determine if the request can be satisfied internally. If so, a structured report is displayed to the user. If not, the request is returned to the knowledge-based system, where a search is conducted of available external sources.

Both on-line data bases and electronic bulletin boards are considered valid knowledge sources, and the knowledge-based system uses inference logic to determine which potential knowledge sources should be accessed, and in what order. A messaging system accesses each knowledge source by drawing on a library of command files that contain the correct communications and access procedures. Dialog is currently limited to search queries and downloading reports.

The downloaded information must be evaluated by the user, who can adjust the search criteria accordingly. If too much information was obtained, the search criteria for subsequent requests of a similar nature could be narrowed. If the needed information was not found, the criteria could be broadened.

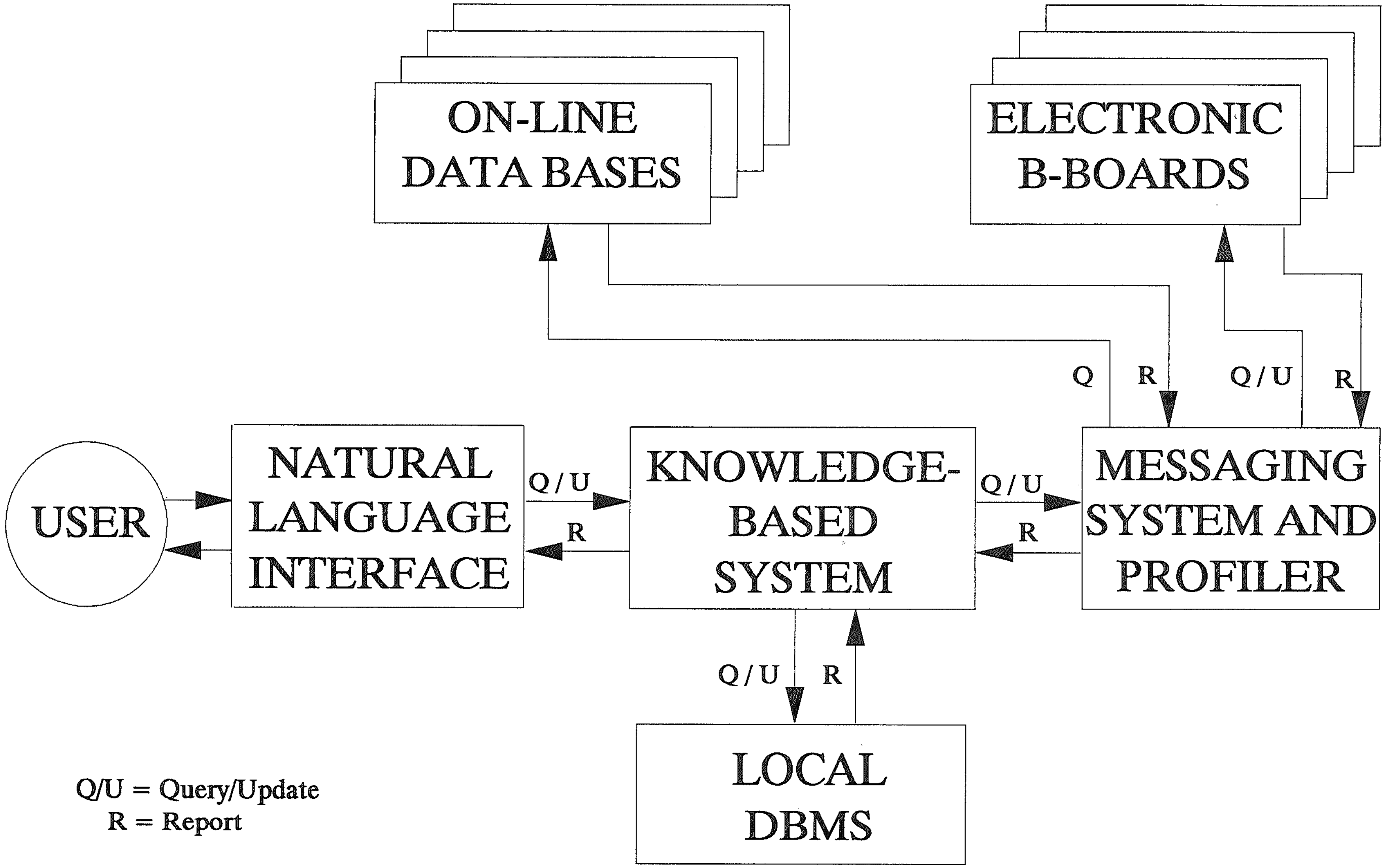


Figure 2. Components of a Semi-Autonomous Network Interface

The knowledge harvester also contains a profiler, in which the user can generate and maintain a weighted hierarchical profile of topics of interest. The system can use this profile to perform periodic searches for both knowledge acquisition and dissemination purposes.

FIRST FRUITS OF THE HARVEST: A VARIETY OF HYBRID CROPS

A unix workstation that implements the architecture shown in Figure 2 has been completed as a shell, and the knowledge base is in the process of being populated. In order to obtain some preliminary performance data, a PC-based knowledge harvester prototype was built in parallel to the unix development effort. The PC-based knowledge harvester does not have a profiler or a natural language interface. However, it does contain a rule-based expert system that is used as an intelligent interface for five different knowledge sources. The expert system prompts the user for the type of information desired, and generates queries by drawing from a library of command files.

A test was conducted using laboratory sessions in which a small sample of users attempted to perform two different tasks. The tasks were to obtain analysts' opinions on a company's stock and earnings performance, and to plan a travel itinerary by combining airline flight information with weather information.

The test sessions were run both with and without the PC-based prototype knowledge harvester, and the results were compared. The basis for comparison was the time taken, and the number of steps required, to perform a search. Data for both parameters were captured through an automated logging process. Any function or control key entry was considered one step, as was the entry of any string followed by the return key. Any manual lookup of information was also considered as one step. Steps were used in addition to time because it was felt that the number of steps would be relatively constant across different users for the same search task. This would tend to equalize attributes such as keyboard dexterity and other motor skills, which, although not in the scope of this effort, could provide expanded insights for future analysis.

The results for the two sample sessions showed that for first-time searches, the number of steps needed to achieve successful knowledge acquisition by using the knowledge harvester prototype was reduced by almost fifty percent. The average time spent for each task was reduced by almost seventy percent. For repeat searches, the average number of steps per task was reduced by thirty percent and the elapsed time was reduced by fifteen percent.

Similar tests are planned to measure potential improvements to the knowledge dissemination and knowledge base maintenance processes. Preliminary observations indicate that improvements

in knowledge dissemination will be similar to those obtained in knowledge acquisition. However, little improvement is expected in knowledge base maintenance, since most of the knowledge base has to be updated manually. The payoff for the maintenance effort must be realized through improved knowledge acquisition and dissemination. However, the payoff is time perishable, and can even result in a performance penalty if the knowledge base is not kept up-to-date.

In summary, the preliminary tests indicate that savings in time and operating costs are possible through the application of an electronic knowledge harvester. Whether savings in investment costs are possible remains to be determined, and will be the subject of future analysis.

INCREASING THE YIELD: GETTING READY FOR THE NEXT SEASON

The preliminary test results point to additional shortcomings that need to be addressed as the development of the knowledge harvester progresses. For example, the intelligent yellow pages, which make up a large portion of the knowledge base, have to be entered and updated manually. In the future, an automated polling process will be used to assist the user in identifying and maintaining a catalog of internetworked resources.

Another shortfall is that both prototype systems are limited to a textual interface. Since knowledge is often visual in nature, a graphical user interface (GUI) that makes use of windowing and multimedia technology needs to be integrated into the system.

New tools are becoming available that will provide improved support for requirements definition and knowledge acquisition (Boose, et al., 1989; Linster, 1989). This improved support will be achieved in part through better problem decomposition and more definitized structuring of requirements and solutions.

Case-based reasoning tools are also emerging that will provide better matching of requirements with potential solutions (Barletta, 1991; Kolodner, 1991; Slade, 1991). Since case-based reasoning approaches are evolutionary, the knowledge base improves over time as case histories accumulate. However, there is a breaking point that is reached when too many cases cause the knowledge base to become brittle. Problems with scaling up are being investigated by developers of very large knowledge-based systems (Silverman, 1991).

Since most of the systems and services on the internetwork deal with the transfer of information, knowledge must be manually inferred, usually through human interaction. A true knowledge harvester must be able to support the machine-mediated transformation of information into knowledge. This can be accomplished through the use of inductive reasoning (McLean, 1991; Parsaye, 1989), and abduction techniques (AbTech, 1991; Punch, 1990). These approaches are not intended to replace human

discovery, but rather to enhance it.

When new knowledge is discovered, the issue remains as to how the knowledge should be encoded in order to support storage, retrieval and dissemination. Knowledge base standards, similar to those used in data base management systems, are needed in order to allow structured knowledge to be shared more easily (Ginsberg, 1991).

Finally, in order for the knowledge base management process to become more autonomous, intelligent interfaces must be placed at the internetwork host nodes as well as at user workstations. Distributed intelligence architectures that would support this capability are in the early stages of development (Murray, 1990; Ortiz, 1990).

REFERENCES

AbTech Corporation, "A Tutorial on AIM and Abductive Modeling," AI Review, 1991, Vol. 4, pp. 7-14.

Barletta, Ralph, "An Introduction to Case-Based Reasoning," AI Expert, August, 1991, Vol. 6, pp. 43-49.

Boose, J. H., D. B. Shema and J. M. Bradsi, "Recent Progress in AQUINAS: A Knowledge Acquisition Workbench," Knowledge Acquisition, June, 1989, Vol. 1, pp. 185-214.

Ginsberg, Matthew L., "Knowledge Interchange Format: The KIF of Death," AI Magazine, Fall, 1991, Vol. 12, pp. 57-63.

Goldberg, David E., Genetic Algorithms, Reading, Addison-Wesely, 1989.

Kolodner, Janet, "Improving Human Decision Making through Case-Based Decision Aiding," AI Magazine, Summer, 1991, Vol. 12, pp. 52-68.

Linster, M., "Towards a Second Generation Knowledge Acquisition Tool," Knowledge Acquisition, June, 1989, Vol. 1, pp. 163-183.

McLean, Stuart, and Charlotte Weise, "DIGRESS, A Deductive Interface to a Relational Database," Journal of the American Society for Information Science, January, 1991, Vol. 42, pp. 49-63.

Murray, Arthur J., "Wide Area Networking Environment," Independent Research and Development Project Description, The McDonnell Douglas Corporation, February 27, 1990.

Naisbitt, John, and Patricia Aburdene, Megatrends 2000, New York, William Morrow and Company, 1990, pp. 304-307.

Ortiz, Charles L. and Marie A. Bienkowski, "Intelligent Information Management in a Distributed Environment," Proceedings of the Fifth Annual AI Systems in Government Conference, IEEE Computer Society Press, 1990, pp. 54-60.

Parsaye, Kamran, et al., Intelligent Databases, New York, John Wiley and Sons, 1989.

Punch III, W. F., et al., "Pierce: A Tool for Experimenting with Abduction," IEEE Expert, October, 1990, Vol. 5, pp. 34-44.

Silverman, Barry, and Arthur Murray, "Full-Sized Knowledge-Based Systems Research Workshop," AI Magazine, January, 1991, Vol. 11, pp. 89-94.

Slade, Stephen, "Case-Based Reasoning: A Research Paradigm," AI Magazine, Spring, 1991, Vol. 12, pp. 42-55.

A Survey of the Use of Wide Area Networks by Librarians

Praveena V. Raman
Massachusetts General Hospital

Candy Schwartz
Simmons College

Abstract

In the present competitive information market, librarians have to offer a variety of value added services, and should also seek to influence the shape of the information society. The formation of a "virtual library" by the Internet, Bitnet, and similar networks has great potential for assisting librarians in achieving these goals. Librarians also stand to benefit from development of the National Research and Education Network (NREN), and indeed are already playing an active role in supporting NREN legislation. In the past few years, numerous meetings have been held, and a wealth of articles and texts published, to help information professionals understand wide area networks and to help them identify ways to use such networks to the best effect. However, it is not clear to what extent and for what purposes librarians are actually making use of networks. With that in mind, this study presents the results of a survey of librarians in the New England area. The purpose of the survey was to discover which networks were being used, by whom, for what reasons, and with what problems. Participants were also asked to identify areas for improvement and further development in wide are networking. Those surveyed represent librarians and other information professionals from corporate, academic, government, research, medical, and public libraries, as well as independent information brokers, consultants, library service suppliers, and educators.

1 Introduction

Advances in technology and the growth in the market for information have made it imperative for the librarians to offer a variety of value-added services and to play an active role in influencing the information society. The appeal of the Internet and similar networks and the potential of NREN is the ability to form a "virtual library" which will aid the librarian in achieving these goals. This has made the networks a popular topic among educators and practicing professionals and a central theme of many conferences and workshops. Much has been written in the professional literature describing the networks and how to use them. A study done in Canada in 1986 (2) examined the reactions and usage of the Netmonth/Bitnet and CDNnet by a general population in post-secondary institutions. However, very little has been published on the nature or amount of usage by the library community. The present study examines the use of networks by librarians in the New England area, and attempts to identify some areas for further research and improvement.

2 Overview

The beginnings of the Internet can be found in 1969, when the Defense Advanced Projects Research Agency (DARPA) funded a project on long-distance packet switching networks. Today the Internet is a mesh of national and international networks connected by means of gateways, and having a consistent form of addressing and common protocols for communication. Up until recently,

access to resources on the Internet was available only to some large universities, laboratories and organizations involved in research and development. The NREN has been proposed to expand and upgrade the Internet. According to the Coalition for the National Research and Education Network (CNREN),

> *The Network will give researchers and students at colleges of all sizes – and at large and small companies – in every state access to the same:*
>
> - *high performance computing tools*
> - *data banks*
> - *supercomputers*
> - *libraries*
> - *specialized research facilities*
> - *educational technologies*
>
> *that are presently available to only a few large universities and laboratories that can afford them. (4, p.297).*

More in-depth information on the Internet and the NREN can be obtained from John Quarterman's *The Matrix* (5) and from LITA's *Library Perspectives on NREN* (3). A good and recent overview on the topic can be found in Lynch and Preston (4).

In 1981, CSNet was established to facilitate collaboration among computer scientists and engineers, while Bitnet was established to facilitate communication among other academicians. In 1989, these two networks merged together. Bitnet supports electronic mail, file transfer functions, and lists, but not interactive sessions. Some other popular networks used by respondents to the survey include ALANet, CompuServe, DialMail, and MCIMail.

Network users have access to two kinds of services: computer mediated communications (CMC) and resource sharing. The most popular example of CMC is electronic mail, a one-to-one communication. Other examples of CMC services are one-to-many (lists or bulletin boards) and many-to-many (conferencing systems). Lists differ from bulletin boards in that users get messages distributed to them as opposed to accessing a host computer to see posted messages. According to Quarterman, a true conferencing system should be able to "display lists of categories and lists of subjects of messages per category, and the user can select messages (either to display or to avoid) by subject, sender, and logical combinations of these and other attributes." In bulletin boards, on the other hand, "users post messages as if on a physical pegboard and with no real idea of who will read them or reply to them. True conferencing systems are used for detailed threads of discussions within continuous topics, and the participants are usually known to each other." (5, p.14)

Features typically supported by networks include remote interactive login (supported by Telnet) and remote file transfer (supported by the file transfer protocols, or FTP). For networks which do not support interactive file transfer, batch transfer can be accomplished through remote job entry, i.e. sending a sequence of commands via e-mail. This is very useful in networks that do not support remote interactive logins or to reduce tying up of interactive ports on the remote hosts. A good example of this would be the retrieval of gene sequences from Genbank using the FASTA or BLAST programs by e-mail.

3 Survey Methodology

A questionnaire (Appendix A) was developed to serve as a guideline for the interviewer in asking questions to gather information in four major areas: personnel access and usage, types of usage,

Category	% of Participants
Academic Libraries, General	34
Academic Libraries, Special	15
Corporate Libraries	27
Government Libraries	8
Information Brokers	4
Public Libraries	4
Information Utilities and Vendors	4
Law Libraries	2
Library and Information Science Faculty	2

Table 1: Breakdown of Participant Categories

Network	% of Participants
Internet	67
Bitnet	58
Dialmail	42
Compuserve	25
Alanet	12
MCIMail	10
Usenet	2
Other	15

The *Other* category in the above table includes
Sprintmail, Fedlink, DECnet, Well, Gennet, Genie and Prodigy

Table 2: Network Usage by Participants

problems encountered during usage and recommendations for improvement and research.

Fifty two New England area librarians were interviewed, either by telephone or in person. The sample included members of the following constituencies: general academic libraries, special academic libraries, medical libraries, law libraries, public libraries, corporate libraries, library and information science faculty, government libraries, information brokers, information utilities and vendors. The breakdown of survey participants into these categories is presented in Table 1.

4 Survey Results

4.1 Networks used

The type of network used depended on where the survey participants worked and to what networks their institutions had access. All the academic and some corporate librarians had access to either the Internet or Bitnet. However, a vast majority of the corporate librarians accessed one of the commercial networks. Table 2 lists the networks with the breakdown in percentages of use.

Category	Participant Response (Percentages)					
	All	More than half	Less than half	Vast Majority	Don't know	N/A
Library Personnel Access	82	4	6	–	–	8
Library Personnel Usage	23	27	42	–	–	8
Other Personnel Access	48	10	11	10	11	10
Other Personnel Usage	17	17	13	12	31	10

Access means who can have an account
Usage means who actually uses the network

Table 3: Personnel Access and Usage

4.2 Personnel Access and Usage

Nearly all the librarians who were interviewed had a good idea as to who in the library had access to the networks and who used them. But few were in the position to give a clear picture of the organization as a whole. Of the participating libraries, 82% did not differentiate between professionals and non-professionals for access to the networks. However, in only 23% of the libraries did all staff use the network. It is common in academic institutions for all faculty and staff to have access to the network, but this says nothing about use. A more detailed breakdown is given in Table 3.

4.3 Exposure to the networks

Workshops, training seminars, staff development programs and conferences were how 50% of the participants had learnt of the existence of the networks and about 29% had read about them in articles, newsletters and technical bulletins. The remaining 21% of the participants had been exposed to networks in a variety of ways: through colleagues, job duties, vendors, professional organizations, corporate or institutional communications and so on. It was interesting to note that one of the participants had been exposed to networks through business cards, which bears out the suggestion of Gurd and Picot (2) that including network addresses in business cards could be a way of proliferating network communication.

4.4 Use of E-mail

88% of the respondents used e-mail, and 56% of the total respondents communicated with both clients and other library professionals. Half of the respondents who used e-mail did not make use of pre-editing and uploading capabilities, preferring to enter messages directly while online. This was principally due to lack of mastery of all the functions of the specific telecommunications package in use. The management of e-mail messages (with respect to printing and archiving) varied according to the nature of the message. Table 4 gives some statistics of the usage and management of the electronic mail function of the networks.

Response	Percentage of participants responding
E-mail usage	
- used	88
- not used	12
Communicated with	
- clients and library professionals	56
- clients only	4
- library professionals only	28
Typed messages	
- online	50
- offline	19
- both	19
E-mail management	
- Printed, stored, downloaded, deleted all messages	37
- Printed all messages	13
- Stored all messages	4
- Downloaded all messages	6
- Deleted all messages	12
- Printed relevant and deleted the rest	6
- Stored relevant and deleted the rest	4
- Downloaded relevant and deleted the rest	2
- Downloaded some and printed the rest	2
- Stored some and downloaded the rest	2
Lists	
- Subscribed to lists	58
- Owners of lists	4

Table 4: Usage and Management of E-mail Function

Fuction	% of participants using function
Bulletin boards on commercial networks	44
FTP feature on the Internet	44
Telnet interactive login	44
Telnet function to use library catalogs	42

Table 5: Use of Functions Other than E-mail

Usage	% of participants
Read and broadcasted	61
Read	31
Neither	8

Table 6: Lists and Bulletin board usage

4.5 Other Uses of the Networks

Those participants who were connected to the Internet used a variety of network functions, such as transferring files from remote host by remote login and logging on to remote hosts and using them interactively. Those participants who only had access to Bitnet used all the above functions but could not logon to remote hosts. Those who did not have access to Bitnet or the Internet but subscribed to one or another commercial networks used a variety of bulletin boards and other functions of their respective networks. Appendix B gives a list of lists and bulletin boards used by the surveyed librarians. Table 5 gives some additional figures for people using functions other than e-mail.

4.6 Use of Lists and Bulletin Boards

58% of the respondents subscribed to one or more lists (interestingly, two were list owners), and 44% accessed one or more bulletin boards through commercial networks (Table 6). Appendix B shows the range of lists and bulletin boards. Of those who subscribed to lists or accessed bulletin boards, 61% read and broadcast, and the remainder were what are commonly called "lurkers" (i.e., they only read messages).

4.7 Remote Access of Text Files and Software Other than Library Catalogs

Only 25% of the participants accessed text files, software, or other remote resources other than library catalogs. Some examples of these resources are listed in Appendix C.

4.8 Problems Encountered on the System

Even though there is a huge number of networks (commercial and non-commercial) connected together, problems due to lost messages or delays in transmitting or receiving them and system downtime did not seem to affect the survey participants. What they found frustrating was not knowing what sort of information was out there, or not knowing how to obtain information about

Problems	% of survey participants
Time delay	
- Experienced delays	38
- No perceptible delays	62
System downtime	
- Experienced downtime	21
- No disruptive downtimes	79
Uploading and downloading information	
- Had problems	23
- No problems	60
- Never tried	17
Directory	
- Felt the absence of directory	59
- Did not feel absence of directory	29
- Not applicable	12

Table 7: Problems Encountered

the different lists and bulletin boards that are already in existence. It was noted that most of the institutions gave some training to the novice users but none to those who were not novices and had not yet become experts in different systems and software. Those with an adventurous spirit fared well and enjoyed the experience, but the majority felt frustrated. Table 7 gives statistics of some of the problems encountered.

4.9 Improvements

The participants were asked what improvements they would like to see to make it easier for them to use the networks. Some of the improvements suggested came directly from the problems that they had faced, which were addressed in the previous section. 59% felt a need for a white pages directory to facilitate easier communication. A very small percentage (8%) of the participants were completely satisfied with their environment and felt that no improvements were necessary. A need for more user training and system support was felt by 33% of the participants. This need was also documented by Gurd and Picot (2) in their paper in 1986 and the need still exists in 1991. Some of the other improvements suggested were: better documentation giving a clearer picture of the ways to navigate the networks, making delivery of electronic messages more reliable, and standardizing the addressing format across the networks. Again, Gurd and Picot (2) had perceived a need for more transparent access to networks and in the same vein, the participants in this study still feel a need for more menus and front ends to make communications less frustrating. There is also a clear need for software that makes downloading and uploading of information easy for even a novice to do by him or herself, as there is a shortage of technical assistance. In some instances, a shortage of hardware also posed a problem in accessing networks. A number of respondents felt that the format of electronic messages received through the Internet or Bitnet should be changed, and that routing information should be compressed and placed in the end of the messages.

5 Future Uses

The participants were asked two additional questions in conclusion of the interview:

1. what other features of the networks would they like to use, and

2. what features would they like to see in the future.

For the first part of the question, 33% of the librarians felt that they did not know enough of what was there on the system to start with, and hence declined to answer the question. Participants who only had access to commercial networks said they definitely would like to have access to all the features of the Internet, a feeling that was also shared by Bitnet users. Some of the Internet users were novices and had not started using the FTP protocol to transfer files and were hoping to do that with some assistance; some of the more experienced users wanted to increase their subscriptions to the lists with the availability of some time. A small percentage, 6%, of the participants wanted to start using some of the shareware software (but were not aware of the possibility of their system being infected by a virus through that channel).

The second question of what they would like to see in the networks in the future was answered with more enthusiasm. Some of the improvements that were suggested in the previous sections were repeated (for example, the need for a directory of addresses, standardization across the networks, and front ends and menus). Other suggestions included information maps, availability of image files and facsimile transmissions, access to networked CD-ROM databases, equitable access to the networks and making the Internet more of a citizen's network. Respondents also listed a number of resources that they would like to have easily accessible in electronic form. These included technical reports from the government and various institutes, IEEE standards, and software that can handle large FTPs quickly.

6 Conclusions

The excitement of having access to a vast amount of information and quick communication was certainly high among the participants. During the course of the survey, some of the beginning network users were introduced to more experienced users for consultation on difficulties. It was noted that librarians need a little more training in the efficient management of electronic mail messages. For example, storing all messages received on the mainframe system uses a lot of space unnecessarily. The participants who were doing a selective combination of printing, deleting, storing and downloading were perhaps the most efficient in the use of system space. It was also noted that though participants pointed out the existence of a huge number of lists and bulletin boards available for subscription, and although they bemoaned the amount of time needed to wade through the messages, not many in group situations banded together to monitor different lists and bulletin boards and exchange pertinent ideas with each other, thus reducing the amount of time needed for individual subscription. Also, in some cases, more than one staff member in the same library subscribed to the same list and did not explore whether a single subscription could be done with multiple staff members accessing in one account instead of many getting the same messages in different e-mail accounts. A greater effort has to be made to have lists and bulletin boards that are refereed so that unncessary noise can be cut and only relevant information can be broadcasted.

Many participants felt the need for online help with technical information and network usage. For all those who are not aware of the existence of the Help-Net list (address listserv@templvm.bitnet), it is highly recommended. It is very helpful for learning the intricacies of the Internet and Bitnet and experienced users give free consultation and technical guidance.

In concluding, it is hoped that as librarians get more proficient in their use of the networks, they would go beyond using it for professional development and faster communications to developing new services for their clients by accessing factual databanks (e.g., gene sequence databases in the medical sciences) to retrieve actual information rather than just the bibliography. Also, they should take the lead in training and encouraging people outside their profession to start using the networks and become comfortable with them.

This study points to the need for several more extensive research projects. One should be a survey of a sample of librarians across the country to see whether some of their needs (such as training) differ in different geographic regions. Another might look in more depth at librarians in specific types of settings (several surveys of this nature are currently in process). A third (and much more difficult) project might study library users to see whether their information needs and information seeking habits change in a networked environment. If that is the case (as one might expect), then librarians must be aware of these changes so that they can alter their services accordingly.

For now, the prinicpal use of networks by librarians appears to be directed to information sharing with other professionals, or to activities related to professional development. It is hoped that as librarians become more "network-literate", they will move beyond professional networking to developing new services for clients. Services which are available now and which are client directed include accessing factual databanks (e.g., gene sequence files, or the Medieval and Early Modern Databank); providing current awareness from lists, bulletin boards, and electronic journals; and, staying abreast of new software and information resources which might serve client needs. In anticipation of NREN or a similar initiative, librarians should also take the lead in training client communities of all kinds, and encouraging them to make use of this vast web of storehouses.

APPENDICES

A Survey Questionnaire

(A) ORGANIZATION: NAME
TYPE

(B) NETWORKS USED:

(1)	ALANET	(2)	BITNET	(3)	COMPUSERVE
(4)	CSNET	(5)	DIALMAIL	(6)	INTERNET
(7)	MCI MAIL	(8)	USENET	(9)	OTHER

Who in the library has access to networks?

Who in the library uses the networks?

Who in the institution, outside the library department, has access to the networks? and who uses them?

How did you find out about the networks?

(C) USES:

Do you use the network for e-mail?

Do you use the e-mail function to communicate with clients and other library professionals?

Do you type the messages offline and then send them or do you create them online?

Do you store the messages, print them or do you download them to a file?

Do you use the networks for any function other than e-mail? If so, what are they?

Do you access bulletin boards and conferencing systems?

If so, which ones?

Do you read and broadcast through the bulletin boards?

Do you send and receive files (other than messages)?

Do you access mainframe supported text files (other than your own)?

If so, which ones?

Do you access mainframe supported software (other than your own)?

(D) PROBLEMS ENCOUNTERED:

(1) Time delays in sending and receiving files and messages
(2) system downtime
(3) uploading & downloading information
(4) abscence of a directory of addresses for communication
(5) How can the system be enhanced and improved at your end?

(F) FUTURE USES:

Are there any features that you do not currently use but would like to?

Are there any features that are not available on the system but you would like to have in the future?

B List of Lists and Bulletin Boards

The following lists and bulletin boards are included as cited by the participants, and do not necessarily represent their formal names.

Lists

AFRICAN-NEWS	ARLIS-L	AUTOCAT	BI-L
BUSLIB	CDROM	CDROMLAN	ETHMUS-L
EXLIBRIS	FEMINIST	GOVDOC-L	HYTELNET
ILL-L	INNOPAC	IR-L	LIBADMIN
LIBREF-L	MAPS-L	MEDLIB-L	MLA-L
MORRIS	NETMONTH	NET-FAX	NOTIS-L
PACS-L	PAM-NET	PRICES-L	SERIALST

Bulletin Boards

America-on-line	Beer
Book Groups	Boston Computer Society
Census Depositories	Economic
Educational Uses of Computing	Electronic Reference
Equestrian	GPO Project
Health	IEEE
Information Professional	Journalism
Language Forum	Literary Reviews
Lotus 123	Movies Reviews
NENON Job Bulletin Board	PC Computers
Personnel	Robotics
Scientists & Engineers	Scifraud
Silent Twister	SLA Employment
Teachers of Foreign Languages	Telecom Div. of SLA
Twin Peaks	US Supreme Court Opinions
Vegetarianism	Zenith Laptop

C Examples of Text Files and Software Accessed on Remote Hosts

The text files and software included in the following list are as cited by the participants, and do not necessarily represent their formal names.

Supreme court rulings and text decisions
Federal databases
Chemical abstracts (STN)
Catalog with Asian studies, mostly Chinese and Japanese
SRI files
CARL book reviews online
Repositories of electronic texts
NREN (Apple computer)
Videotext

1032WP
Statistical packages
Mas graphics
Financial budgeting software
Grants management system
PRISM
EPIC

REFERENCES

1. Arms, Caroline R. "A New Information Infrastructure." *Online.* 14(5)15-22; 1990.

2. Gurd, G.; Picot, J. "A Study of Atlantic Canadian User Reactions to Two Inter-university Electronic Networks." *The Canadian Journal of Information Science.* 11(134)102-121; 1986.

3. *Library Perspectives on NREN: the National Research and Eduction Network.* Chicago, IL: Library and Information Technology Association. 1990.

4. Lynch, Clifford A.; Preston, C. M. "Internet Access to Information Resources." *Annual Review of Information Science and Technology (ARIST).* 25()263-311; 1990.

5. Quarterman, John S. *The Matrix: Computer Networks and Conferencing Systems Worldwide.* Bedford, MA: Digital Press. 1990.

BACKGROUND READING

- Arms, Caroline R. "Using the National Networks: Bitnet and the Internet." *Online.* 14(5)24-29; 1990.

- Brett, G. H. II. "Online Resources: Getting There From Here." *Educom Review.* Fall 1990, pp 48-53.

- Cisler, Steve. "The Library Community and the National Research and Education Network." *Wilson Library Bulletin.* June 1990, pp 51-55.

- Cline, Nancy. "Information Resources and the National Network." *Educom Review.* Summer 1990, pp 30-34.

- "Communications, Computers and Networks." *Scientific American.* Special Issue, Sept 1991.

- Delfino, Erik. "Email Connections: It's Still a Jungle Out There ... But It Is Getting Better." *Online.* 14(5)31-35; 1990.

- Landweber, L. H.; Jennings D. M.; Fuchs, I. "Research Computer Networks and Their Interconnections." *IEEE Communications Magazine.* 24(6)5-17; 1986.

- Learn, L. L. "Networks: A Review of Their Technology, Architecture and Implementation". *Library HiTech.* Issue 22, 1988, pp 19-49.

- Lynch, C. A. "Linking Library Automation Systems in the Internet: Functional Requirements, Planning and Policy Issues." *Library HiTech.* Issue 28, 1989, pp 7-18.

- Martin, W. J. "The Network Marketplace and Its Implications." *Aslib Proceedings.* 42(6)165-171; 1990.
- Nielsen, B. "Finding it on the Internet: The Next Challenge for Librarianship." *Database.* Oct 1990, pp 105-107.
- Perry, D. G.; Blumenthal, S. H.; Hinden, R. M. "The Arpanet and the Darpa Internet." *Library HiTech.* Issue 22, 1988, pp 51-62.
- Raeder, A. W.; Andrews, K. L. "Searching Library Catalogs on the Internet: A Survey." *Database Searcher.* Sept 1990. pp 16-31.
- Stahl, N. J. "Using the Internet to Access CARL and Other Electronic Information Systems". *Science and Technology Libraries.* 11(1)19-29; 1990.
- Woodrow, R. B. "Telecommunications and Information Networks: Growing International Tensions and Their Underlying Causes." *The Information Society.* 6()117-125; 1989.

Part II

Panel Sessions

Session on:

Libraries and National Networks

Merri Beth Lavagnino
Assistant to Head, Systems
Yale University Libraries

Paul Evan Peters
Director
Coalition for Networked Information

Carol Parkhurst
Assistant University Librarian
for Systems and Technical Services
University of Nevada

Peggy Seiden
Head Librarian
Pennsylvania State University at New Kensington

Speakers will discuss what libraries should and are doing about the information available over national networks. Attendees will learn some basic key words, commands, and categories of networked information, to find out what is available and how to locate it. Examples of projects libraries have implemented to provide access to these sources will be presented.

Federal Natural Resources and Energy Programs

Nancy Y. McGovern
Center for Electronic Records
National Archives and Records Administration

Jeanne Young
Archivist
National Archives and Records Administration

Frank Splendoria
Bureau of Land Management

Judy Drumm
Department of Energy

A mid-year meeting in New Mexico provides an opportunity to get a look at what technological applications the Federal Government is exploring in the areas of energy and natural resources. The three Federal agencies listed have very active regional programs and a variety of interesting projects to discuss. The focus of the session will be work stations and user interfaces using case studies from the agencies. For example, the Bureau of Land Management has the Texas Acquired Minerals Project (TAMPS) which involves the conversion of a large, significant system from manual to electronic form with consideration of access to the information, formats, GIS considerations, federal and state cooperation, etc. The Department of Energy has a large network of contractors to be coordinated as part of the Information Resources Management program dealing with records manaagement, and information flow.

Session on:

State and Regional Networks

Michael Lynch
Systems Librarian
Bucknell University

Thomas Bajzek
Executive Director
Pennsylvania Research and Economic Partnership Network

Ward E. Shaw
Executive Director
CARL Systems, Inc.

Jeff Ogden
Associate Director
Merit Network

Speakers will address the roles of state and regional networks. What do they do? How do they relate to the national network: By supporting and expanding it? By providing parts of it in terms of equipment and resources? They will also describe specific projects currently being carried out by these networks.

Multimedia Issues in Networks

Karen Kaye (Moderator)
Multimedia Initiative Project Manager
NASA Scientific and Technical Information Program

Bob Conley
Computational Services
Air Force Space Technology Center
Kirtland Air Force Base

Barbara Baker
Starlight Networks

Kathleen Burnett
Rutgers University

Bob Conley will discuss data compression, a multimedia networking enabling technology, in terms of a computer program developed at Kirtland that compresses color graphics animation sequences for local storage and transmission to remotely networked sites in a significantly reduced size. Following transmission, playback of the animation can be accomplished on a wide variety of workstations ranging from personal computer class machines to high and scientific and engineering workstations.

Barbara Baker will provide a technology primer for those planning to network multimedia information at their sites. The emphasis in the presentation will be on the issues one needs to consider to network video on LANs successfully. The presentation will be general in nature and will not include product specifics.

Kathleen Burnett will present her paper "Multimedia as Rhizome: Design Issues in a Network Environment," which will extend the analogy of the root structure of a rhizome to that of the infrastructure of large networks such as the Internet. The variability we see in multimedia, both from implementation to implementation, and within single applications, is not unlike the variations we see in rhizome growth – from planting to planting, and within a single growth.

Session on:

Telecommuting in an Information Environment

Robert Gresehover, Moderator
Johns Hopkins Applied Physics Laboratory

Edmond J. Sawyer
Consultant
The JELEM Company

Jessica L. Milstead
Principal
The JELEM Company

This session will address the growing interest in using computer and telecommunications technologies to work at home, in a satellite office, or at a customer's worksite. Telecommunications companies are developing systems that facilitate this changing structure of the workplace through automatic routing of voice and data to company service employees working outide the office.

NEEDS (The National Engineering Education Delivery System): If we build it (according to standards)they will come!

by John M. Saylor, Director, Engineering Library, Cornell University, Ithaca, NY 14853
e-mail: John_Saylor @qmrelay.mail.cornell.edu

The NSF is providing funds for coalitions of engineering educational institutions to improve the quality of undergraduate engineering education. A hypothesis that we are testing is that people can learn better in environments that allow self-paced and/or collaborative learning. These environments need to provide information in a full range of formats including not only the traditional blackboard and lecture but also interactive software modules, video segments, pictures and graphics, outlines and text. The main tools for providing this environment are called incorporated in NEEDS. NEEDS includes a fully networked distributed multimedia database for storing, searching, and retrieving this information, electronic classrooms for learning and teaching with the information, and authoring studios where the information is massaged into modules for instruction. We are initially building these tools for use by students and instructors in engineering education. Eventually, these tools will be used by instructors and students at the other two ends of the education spectrum, K-12 and continuuing education. The theory is that if we are succesfull in building effective tools, we will attract and retain a greater diversity and number of young students, especially women and underrepresented minorites to the engineering profession. The recent federal committment to the National Research and Education Network will provide the networked electronic infrastructure on which to build NEEDS and help accomplish a major node in the vision of a Digital Library System (CERF).

Project Goals

Officially this project was born on September 30, 1990, when the NSF funded its first two Engineering Education Coalitions. The goals of the NSF program are: (1) to increase dramatically the quality of undergraduate engineering education as well as the number of engineering baccalaureate degrees awarded, especially to women and underrepresented minorities; (2) to design, implement, evaluate and disseminate new structures and fresh approaches affecting all aspects of undergraduate engineering education including both curriculum content and significant new instructional delivery systems; and (3) to create significant intellectual exchange and substantive resource linkages among major engineering baccalaureate-producing institutions and other major and smaller institutions.

The major focus of this project is to restructure and provide tools for engineering education-- not soley engineering research. The products of this work will be multimedia modules designed to enhance learning for all engineering students regardless of gender, ethnicity, or race. These multimedia modules will consist of the full range of graphical materilas (interactive software modules, video segments, pictures and graphics, outlines and text). (LYNCH)

What is Synthesis

The name Synthesis comes from the group's overall theme of interdisciplinary, multilevel (pre-college through postgraduate) integration of engineering knowledge, including design. This "synthesis" involves putting together a structure of individual parts (curriculum, supporting technologies, recruitment and retention, and linkages) to make up a complex whole. The Curriculum component involves projects designed to revitalize the engineering curriculum both through innovative instructional modules and through systemic endeavors. Supporting Technologies contains projects that provide the supporting technology needed to accomplish this curricular change. Recruitment and retention (known in Synthesis parlance as Pipeline) refer to the need to attract and retain historically under-represented groups to engineering as a profession. Linkage refers to marketing methodologies used to promote the value and attractiveness of engineering as a profession beyond the traditional classroom to the public through high impact channels such as professional societies, television, and advertising. All Coalition projects (75 in all) are collaborative in nature, designed for dissemination to engineering schools throughout the country as well as K-12 levels.

NSF received 10 proposals, from teams involving 104 institutions. Two groups of universities received funds of $15 million for five years -- the Synthesis Coalition(the subject of this paper) and the Engineering Coalition of Schools for Excellence in Education (ECSEL). The Synthesis Coalition schools are California State University at San Luis Obispo, the University of California at Berkeley, Cornell University, Hampton University, Iowa State University, Southern University, Stanford University, and Tuskegee University. This coalition represents diversity in geographical locations as well as variety in size, mission, and institutional type.

For a more in depth introductory look at both Coalitions I recommend the following. The first is: "Synthesis: A Coalition Approach," by Robert J. Thomas, Professor of Electrical Engineering at Cornell University and Director of Cornell's Synthesis Coalition projects. (*ASEE PRISM*, pp.14-16, Preview Issue, 1991.) The second, which gives an introduction to the ECSEL Coalition is: "Engineering Coalitions Find Strength in Unity" by Jeff Meade. (*ASEE*

PRISM, pp. 24-26. September, 1991). For more detail about curriculum reform activties in Synthesis, see "Refreshing curricula" by George Watson in *IEEE Spectrum,* March 1992, pp 31-35(.

Activities to Date

Curriculum Projects

Interdisciplinary Multimedia Case Studies

Case studies are prepared with computers and hypermedia. Students can navigate at their own speed, via workstations, through databases to learn not only scientific and technological background to the case study but also the social, historical, business, and environmental implications related to the case. One project that illustrates this idea is titled "Collaborative Design in a Networked Multimedia Environment." This project has recently been discussed in *EDUCOM Review* , Volume 27, number 1, 1992, pages 31-33, ("Collaborative Design in a Networked Multimedia Environment, " Gay, G.K. and Thomas, R.J.) and in *CD-ROM Professional,* Volume 5, Number 2, 1992, ("Joining Digital Hypermedia and Networking for Collaboration in Engineering Design: a project's early consideration," Mazur, F.E. and Gay, G. K.). The design objectives of this project are twofold. First the learning environment is to be patterned after real-world employment. In industry, concurrent engineering (CE) principles are applied to solving design problems. Using LAN's and extensive networks, engineers work together with representatives of purchasing, marketing and others in the company during product design and review. CE is not very common in engineering education. As a result, graduates of engineering schools are not adequately prepared for working in this way when thay are newly employed by industry. The second design objective is that contextual influences are to be emphasized. Critics from industry say that present day teaching is too simplified and frequently lacks depth of knowledge from other fields of study that relate to a problem. A key tool in testing these objects is the hypermedia database (NEEDS), which I will discuss later. According to the aforementioned article (Mazur and Gay) " The information nodes in the database are to be contextually rich and expressive in the presentation of content and are to reflect multiple representations so that students can better comprehend key engineering design concepts and principles."

Recruitment Projects

Beginners can learn design methodologies even though they lack background in engineering fundamentals. In this regard, freshman engineers are assigned simple but practical design

projects to involve them in real world projects and keep their interest piqued from the beginneing. One such course is the Spatial reasoning course.

Linkage Projects

Supporting Technologies Projects

NEEDS

As mentioned above, a cornerstone of the project is the National Engineering Education Delivery System (NEEDS). NEEDS consists of: (a) multimedia databases of curricular materials consisting of data elements ranging from simple text to full motion video, which are connected to; (b) courseware development studios for faculty and; (c)high-technology classrooms connected through high speed networks, both on campus and internationally through the NREN. In NEEDS we are building a major digital library node in the networked Digital Library System.

Standards Study Project (SSP)

The Standards Study Project is a separately funded, five year project whose purpose is to identify the technologies required for NEEDS; to identify the problem areas due to lack of standards in information storage, retrieval, transfer and manipulation; to identify the existing and developing relevant standards; and to suggest effective courses of action to allow NEEDS to develop in concert with emerging standards and technologies. We have convened a Standards Study Advisory Group of experts from industry and academe and have held three national meetings our first year. I described these meetings in more depth in the premier issue of the electronic publication *Issues In Science And Technology Librarianship*, (December 1991.)

Accomplishments of NEEDS and SSP

The Standards Study Project has evolved to become the leading force in Synthesis in planning and developing NEEDS. Planning the architecture of the database, electronic classrooms, and courseware studios has been the focus of our first two major meetings.

Meeting of October 1991

The Standards Study Project Advisory group is made up of Synthesis Coalition members as well as industrial partners from John Wiley, Bellcore, The Interactive Multimedia Association, IBM, Mitre, Hewlett-Packard and others. This group divided into three committees based on the database standards, the electronic classroom , and the courseware studio. In this first meeting the Database Standards Committee recommended that:

(1)set up a coalition-wide editorial board to deal with policy and planning for the NEEDS database;

(2) more precisely define access control requirements for the NEEDS repository;

(3) define in detail the functional requirements for the institutional NEEDs servers and differentiate the functions of the central repository and the institutional sytems more precisely.

We defined a preliminary two part architecture for the central NEEDS database. The first part would be a repository of source material. The second part would be a catalog of MARC-like records describing the source material. The central NEEDS database would be a repository of the full spectrum of source material (simple text to multimedia modules) and be accessible via the internet. The database would be built on the MARC record and would support the Z39.50 Search and Retrieval protocol. (LYNCH)

On of the goals of this group is to provide a more formal vision of the overall system architecture of the database for the coalition to react to and discuss.

Meeting of March 1992 - Will be presented at ASIS Midyear, May, 1992.

Dificulties Encountered

• Great volume of material being produced with no way to organize it

Currently there are Synthesis Coalition courses being taught and designed. A vast quantity of graphical materials such as slides, video segments, graphics are being produced but no database exists as yet in which to sore this material. This is happening on a large scale.

• How to make this material available to and usable by others

There has been no firm committment or plan on how this material will be cataloged or indexed.

• Lack of adopted Multimedia Standards

Content material (slides, video, etc) are being stored two ways compressed in Quicktime (not an industry standard) and as uncompressed images in order to provide access for use by those with differing equipment resources.

•Tenure Issues

Tenure is largely based on research publication. Production of materials for undergraduate education is not currently mainstrained in the reward structure for scholarly promotion

•Time involved in creation of multimedia materials

Will already overburdened faculty and staff take the time to learn new techniques and skills to create and perform in the multimedia environment.

•Intellectual Property Rights issues

This is really the dam that is holding back the flood of availabilty and use of electronic multimedia information. It is one of the two major impediments to the large scale conversionof paper collections to electronic form, the other being cost. ((8)LYNCH, 1991)

Questions to be Answered

Do we collect in db everything being produced?
At this point the scenario is that when a project produces a deliverable that the investigator is happy with, we will store it.

If not, who reviews what goes in?
The review board issue has not been resolved.

We will build what we can afford or are given
At this point it is felt that we have (or will have)unlimited storage so we will probably store everything produced by the Synthesis projects. It is believed by some that in the near future, "Malthusisan concerns about data overpopulation are easily solved by a combination of advances in high density storage systems and techniques which allow data to die a natural death. (Kahn, Robert E & Cerf, Vinton G. , The Digital Library Project Volume 1: The World of Knowbots (Draft). Corporation for National Research Initiatives, 1988, page 10)

Library's role

The business of libraries is to select, organize, preserve, and provide access to information. The bussiness of managing multimedia education databases in this way brings up many issues and unanswered questions.

Finding Relevent Material

"Finding relevent material, and even learning of its existence, is often a massive challenge. This problem is not unique to the research world domain. It plagues virtually every information-dependent human endeavor." - (Kahn, Robert E & Cerf, Vinton G. , The Digital Library Project Volume 1: The World of Knowbots (Draft). Corporation for National Research Initiatives, 1988, page 14)

Selecting Material

Does our role in the Digital Library System diminish in terms of selecting materials? Who will decided what goes into the multimedia educational database?

Preserving Material

What role do we play in preserving electronic, multimedia, educational material? What new infrastructure and expertise do we have to develop in order to archive this material.

Providing Access

Providing access to multimedia objects via electronic networks is currently the subject of much research. The only off-the-shelf library software currently available to handle this is provided by VTLS Inc. in their VTLS InfoStation front end to their integrated multimedia OPAC system. The problem with VTLS at this point is that it only runs on the proprietary NeXT harware and software. (LEE)

Conclusion

The business of education is to create, transmit, store, retrieve, display, manipulate, and interact with information

The business of libraries is to select, organize, preserve, and provide access to information.

Computers and their networks are tools used to facilitate these activities.

Ones judgement cannot be better than the information upon which it is based. Given the truth, one may still go wrong when one has the chance to be right. But not given news or presented only with distorted and incomplete data, you destroy the whole reasoning process... from Arthur Hays Sulzberger , address, NYS Publishers Association (8/30/48)

NEEDS is a crucial tool that the engineering educators in Synthesis will rely on in their business of educational reform. Librarians will assist in the process of building NEEDS based on their long term experience and knowledge of selecting, organizing, preserving, and providing access to information. Network computers and their associated databases of information and applications for manipulation will provide the tools for facilitating the work. This is truly a collaborative project.

References

1) Watson, George F. (1992) "Refreshing Curricula", IEEE Spectrum, vol , no. March, pages 31-35

2) Kahn, Robert E & Cerf, Vinton G.(1988) , The Digital Library Project Volume 1: The World of Knowbots (Draft). Corporation for National Research Initiatives

3) Thomas, Robert J, (1991) "Synthesis: A Coalition Approach," *ASEE PRISM*, pp.14-16, Preview Issue, 1991.)

4)Meade, Jeff "Engineering Coalitions Find Strength in Unity" *ASEE PRISM*, pp. 24-26. September, 1991

5)Gay, G.K. and Thomas, R.J. (1992) "Collaborative Design in a Networked Multimedia Environment." *EDUCOM Review* , Volume 27, number 1, 1992, pages 31-33,

6) Mazur, F.E. and Gay, G. K. (1992)"Joining Digital Hypermedia and Networking for Collaboration in Engineering Design: a project's early consideration," *CD-ROM Professional*, Volume 5, Number 2, 1992,

7) Lynch, Clifford, A. (1991) " The Technologies of Electronic Imaging," *Journal of the American Society for Informnation Science,* 42(8):578-585

8.) Lynch, Clifford A. (1991) " The Development of Electronic Publishing and Digital," *Electronic Networking,* 1(2):6-22

9.) Saylor, John M. (1991), " ", *Issues in Science and Technology Librarianship*

Session on:

Progress Towards Remote Image Serving: Case Studies in the Arts and Humanities

Joseph Busch, Moderator
Systems Project Manager
Getty Art History Information Program

Thomas E. Dackow
Gail Egan
Angela Giral

As the technology and infrastructure to document, capture, and transmit images has developed, archives, museums, and galleries have begun to take advantage of this situation to develop the ingredients and integrate them into systems for remote image serving. This session will present case studies which illustrate the evolution of remote image serving through projects to coordinate the documentation of objects, image capture projects, and the integration of documentation and objects into systems and products. If feasible, the session will include demonstrations of the facilities provided by each project either during the presentation or in a separate venue at the conference, as appropriate.

Presenters:

Canadian Heritage Information Network, Communications Canada, 365 Laurier Avenue, West, Ottawa, ONT K1A 0C8, Canada. Tel: 613-992-3333. Fax: 613-952-2318. Not yet confirmed. The basic ingredient for a database of images is the development of documentation about them answering the questions what is it? and where is it? The Canadian Heritage Information Network (CHIN) provides a time shared resource for cataloging museum objects. CHIN is a bibliographic utility style resource which is creating national databases of museum collections in Canada.

AVIADOR Project, Avery Architectural and Fine Arts Library, Columbia University, New York, New York 10027. Tel: 212-854-3501. Fax: 212-749-0397. Internet: giralcunixc.cc.columbia.edu Not yet confirmed. The AVIADOR project has produced documentation quality images of the architectural drawings in Columbia University's Avery Library collection. These images have been transferred to videodisk and cataloged on RLIN using the MARC AMC format. RLIN has just completed development of an interface between the videodisk player and PC terminal which provides automated access to images on the videodisk via the RLIN record. This scheme supports remotely served data and controls a locally replicated image base accessed on a videodisk player.

Q Systems Research Corporation, 75 Avenue of the Americas, New York, NY 10013. Tel: 212-941-1660. Not yet confirmed. Q systems has created a commercial product which provides remote access to documentation and images of objects offered for sale by the major art auction houses. This application melds the ingredients of a data- and image-base into an online commercial product. Q Systems takes advantage of the evolving telecommunications infrastructure and image processing hardware to remotely serve images via satellite links from a centrally maintained database.

Information on the Internet: Discovery, Access, and Use

A Panel Discussion

This panel assembles representative specialists to discuss problems and approaches associated with describing, locating, accessing, and using resources on the Internet. The panel will introduce audience participants to the problems associated with providing systematic access and traditional library services in a global, high-speed network environment.

Presentations by panelists cover a range of current research projects and various approaches to defining the problem and exploring the solutions. Topics include a description of Internet textual resources, the development of directories and directories of directories, indexing in a volatile environment, and automated wide-area information discovery and retrieval.

Comprising a variety of perspectives including librarianship, information retrieval, and database design, this panel will focus on a clear discussion of the problem, a review of current research, and an assessment of the strengths and weaknesses of existing and proposed solutions.

Assessing Information on the Internet

Martin Dillon, Director, Office of Research, OCLC

This paper presents findings of work in progress funded by the U.S. Department of Education, Library Programs. Discussion includes detailed descriptions of the characteristics of textual information available on the Internet, the results of a cataloging initiative to test the applicability of MARC and other formats for file description, and the results of early attempts to automate the cataloging process for Internet files.

Distributed Information Characterization and Search on the Internet

Michael Schwartz, Department of Computer Science, University of Colorado, Boulder

The decentralized nature of the Internet has two broad implications on resource discovery. First, automated means are needed to extract attribute information from resources. Manual classification is painstaking and error-prone, and produces indices that quickly become dated and incomplete. Second, information should be distributed and organized so that it can be searched flexibly, supporting many different "views." Typical hierarchically organized directories provide poor support for such views. In this talk we present a model for resource discovery

called Distributed Two-Phase Search, which supports automated attribute extraction and flexible searches. We discuss the model in the context of a number of research prototypes and studies carried out in the Networked Resource Discovery Project at the University of Colorado, Boulder.

Wide Area Information Servers: A Supercomputer on Every Desk

Brewster Kahle, Project Leader, Wide Area Information Servers, Thinking Machines Corporation

While computers have come to be used by professionals in all fields, finding and accessing information electronically that is not on your local file server has been limited to the trained and tolerant. This project attempts to change this by giving users simple user interfaces for finding servers and accessing the information on them.

Thinking Machines, Apple, Dow Jones, and Peat Marwick companies joined to make an information system for executives that would access personal, corporate, and published information in one easy-to-use interface. This system includes a English language question-answering mechanism, personal newspapers, and remote servers.

This talk will report on this system for electronic publishing and will discuss the set of Internet tools that are being given away to help catalyze a market of information servers.

Session on:

Library Automation and Networking: Issues and Opportunities

Philip Doty, Moderator
School of Information Studies
Syracuse University|

Carol A. Hert
Doctoral candidate
School of Information Studies
Syracuse University

The panel will explore the increasing interconnection between traditional library automation in initiatives and those occurring in networking. Networking initiatives provide opportunities for new roles and services in libraries. These opportunities are not without attendent risks and issue areas such as technology (e.g., security and standardization), planning and management, training and education, financing, and philosophy. Fortunately, librarians' history as service providers and managers of information resources, as well as our 20-year experience with library automation activities, has prepared us to understand the issues involved in networking and also provided us with the skills necessary to manage the implementation and the ongoing use of networked resources. The goal of the panel is to explore the issues and their impact on library setting as well as to explore innovative solutions and efforts in the library community which seek to bring together networking and automation.

American Society for Information Science
SIG/BC

Mid-Year Meeting, May 1992

Session Title: Scientific Information in a Network Environment

Moderator:

Julie M. Hurd
Science Library M/C 234
University of Illinois at Chicago
P.O.Box 8198
Chicago, IL 60680

Panelists:

Ann Bishop
Graduate School of Library and Information Science
University of Illinois
410 David Kinley Hall
1407 West Gregory Drive
Urbana, IL 61801

Peter Liebscher
SURAnet
8400 Baltimore Blvd.
College Park, MD 20740

Harry P. Llull
Centennial Science & Engineering Library
University of New Mexico
Albuquerque, NM 87131

Session Description:

This panel will address issues related to scientific information in a network environment. The panelists include **Ann Bishop**, a faculty member whose research and writing has focused on information organization and retrieval, information-seeking behavior, information policy, scientific communication and electronic networking; **Peter Liebscher**, Manager of the Network Information Center for SURAnet, a National Science Foundation network, whose research examines changes in scientific communication as a result of new electronic communications technologies such as high speed, wide area networks; and **Harry Llull**, Director of the Centennial Science & Engineering Library at University of New Mexico who is active in the Coalition for Networked Information and has recently initiated an electronic publication on scientific and technical libraries for the Science & Technology Section of the Association of College and Research Libraries.

These panelists will discuss the changing nature of scientific communication as increasing amounts of scientific information are made available over networks. The speakers will describe the types of

information on networks such as experimental data, software programs, and bibliographic and reference data, and compare these resources to those available in more traditional, paper-based formats or as computer-readable files on tape or disk. Are there problems in identifying or gaining access to network information? Are there problems of data integrity and equality/ease of access? Are new information policies or infrastructures needed to resolve such problems? What might be the role of NREN, the National Science Foundation and other government agencies? How might professional societies be involved? Can both profit- and not-for-profit organizations interact effectively in this environment? How do information professionals assure full participation in these and future developments?

If scientists take full advantage of the possibilities for enhanced communication over networks, how will established communication patterns be likely to shift? How will information-seeking behavior and scientific research adapt to the accelerated pace of information transfer and to the increasing amounts of information readily available? Are any scientists or groups likely to be information-poor because of sociopolitical or technological attributes they share? Will scientists' reward structures need to alter to reflect the existence of such new developments as computer conferencing and electronic journals? What changes might be anticipated for scientific journals and for scientific information services? Will long-established peer review processes survive in a network setting?

SIG/BC invites active participation in this discussion by all who share an interest in management, dissemination, and communication of scientific information and the networking technologies that promise to accelerate and enhance the research enterprise.

Non-Bibliographic Uses of Z39.50

Margaret Baker
University of California at Berkeley

Though its roots are in the bibliographic world, Z39.50 offers valuable opportunities for other kinds of applications. We chose to use it for an information server we are developing at UC Berkeley and have been working to design and develop a Z39.50 implementation for full text and other non-bibliographic uses. Other UC projects have chosen to use it as well, including the systemwide Sequoia 2000 project and Berkeley's Museum Informatics Project. Last summer, discussions with the developers of CWISP, the Campus-Wide Information Systems Protocol, led to an agreement to incorporate the data structures they needed as Z39.50 record transfer syntaxes. Z39.50 is not the answer to everyone's problems, but its utility extends far beyond the library community.

This paper describes some of the non-bibliographic uses underway, and how these applications are to be implemented.

Advanced Computer and Engineering Research to Serve Medical Information

Session Chair: Dr. George R. Thoma

The National Library of Medicine (NLM) is one of the leaders in the effort to make the goals of the High Performance Computing and Communications Initiative a reality. Components of this initiative include: high performance computing systems; advanced software technology and algorithms; the National Research and Education Network (NREN); and basic research and human resources.

NLM is active through its intramural and extramural programs in several of these areas. Projects include:

- Development of a digital library of the three-dimensional structure of the human body at submillimeter level resolution;
- Creation of an online image archive of digitized radiographs (collected as part of a nationwide study on health and nutrition), and an infrastructure for its access. This involves development of linkages via Internet among sites for high resolution digitizing of these radiographs, sites where they are stored on optical media, and sites where they may be retrieved, edited, and displayed;
- Implementation of remote access via client-server software for sequence similarity searching and text retrieval to support molecular biology research;
- Development of techniques to update, enhance, and edit NLM's Metathesaurus™, a product of UMLS, from remote sites linked to NLM via Internet.

This session offers presentations describing objectives and research activity in these areas.

Dr. Michael J. Ackerman

Visible Human Project

The National Library of Medicine has long been a world leader in the archiving and distribution of the print-based images of biology and medicine. NLM has also been a pioneer in the use of computer systems to encode and distribute textual knowledge of the life sciences. NLM's Long Range Planning effort of 1985-86 foresaw a coming era where NLM's bibliographic and factual database services would be complemented by libraries of digital images, distributed over high speed computer networks and by high capacity physical media.

The NLM Planning Panel on Digital Image Libraries in Biology and Medicine recommended that NLM should undertake the building of a digital image library of volumetric data representing a complete normal adult human male and female cadaver. This "Visible Human Project" would include digital images derived from computerized tomography, magnetic resonance imaging, and photographic images from cryosectioning. This would require the establishment of a working group to establish standards for acquisition, computer representation of the image data, and distribution of the digital library. The "Visible Human Project" will serve as a cornerstone for future sets of related image libraries and as a test platform for developing high performance computing and communication imaging and rendering methods.

Dr. George R. Thoma

Digital Xray Prototype Network (DXPNET)

The overall goal of this project is to establish a radiographic image archive containing the digitized images of 17,000 xrays collected during the second National Health and Nutrition Examination Survey (NHANES), and to provide access to this archive over the INTERNET. The project is a collaboration among the NLM, the National Center for Health Statistics (NCHS), the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS) and the University of California at Los Angeles (UCLA).

The project involves the evaluation of: techniques to index the image database, software mechanisms to link the image collection to the NHANES data set, image quality issues, retrieval and archiving requirements, image compression techniques, and issues related to high speed image communications. It also involves the integration of hardware components that fulfill the requirements of workstations that allow quality control by technicians, the access and evaluation of images by radiologists, and the development of a networked archival optical disk-based storage facility.

At present, a workstation has been completed for the quality control (QC) of xray films

digitized by collaborators at UCLA. The workstation, built at NLM, comprises a high resolution (1K x 1K) display, a WORM drive, an IBM AT compatible controller, and image processing boards and software. NCHS staff members are currently performing QC using this workstation.

The next steps in DXPNET are: to establish an image archive by means of an optical disk jukebox controlled by a UNIX workstation; to develop image communications linkages employing the INTERNET; to develop and deploy prototype workstations for radiologists to access the archive, retrieve the images and develop standardized readings; to develop file management software; to develop image file access software; and to evaluate the system.

Dr. Dennis A. Benson

Network Services for Molecular Biology

The National Center for Biotechnology Information at the NLM designs, maintains, and distributes databases which contain information vital for molecular biology research. The focus of its database activity is the GenInfo system of databases which integrate DNA and protein sequence information with bibliographic and abstract records from MEDLINE. Network services are being developed which will support remote access via client-server software for sequence similarity searching and as well as text retrieval. In order to encourage the independent development of software that can interact with structured biological data, an ISO-standard data description language, ASN.1 (Abstract Syntax Notation No. 1), is being used to define the data objects and the interfaces needed to couple the database to retrieval and analysis software. Prototype network services are now being tested and full operation is planned by October, 1992.

David D. Sherertz

Toward Concurrent Distributed Thesaurus Maintanence and Enhancement

The National Library of Medicine is committed to maintaining an annually updated "Metathesaurus" of biomedicine as part of its Unified Medical Language System (UMLS) initiative. Meta-1.1, the second version of the Metathesaurus, containing information about 64,000 concepts, was alternately edited at the NLM in Bethesda, Maryland, and "computed" by Lexical Technology, Inc. (LTI) in Alameda, California, using the INTERNET. Future (larger) versions of the Metathesaurus will be centrally maintained and controlled by the NLM, but edited by off-site (non-NLM) editors, enhanced by domain experts at distant academic medical centers, and translated by foreign MEDLINE centers. Current plans call for continued incremental evolution of workstation-based, bit-mapped interfaces, supported by a distributed relational database, connected, initially, via the INTERNET and, eventually, by the High Performance Computing and Communications Network. When supplied with planned-for network and computing capabilities, Metathesaurus editors, enhancers and translators will be able to assess the global (Metathesaurus-wide) impact of their decisions and additions in "real" time.

Educating the Networking Information Professional - SIG/ED

In their status report on the NREN McClure *et al.* identified several problems associated with the NREN, including "lack of user friendliness" with potential users dissatisfied with the effort they must expend to acquire networking knowledge; scarce instruction, documentation, and troubleshooting support; and inconsistent format and retrieval processes. They conclude that writings about the NREN display "a surprising lack of concern with the need for user support, education, and training... Until networks become easier to use, many scientists and researchers may be reluctant to expend the time and effort needed to learn how to overcome these obstacles."[1]

How do demands on information professionals change when the resource in use is an unseen, ever-changing, world-wide network of computer-based information? Can today's students be prepared to cope with, even anticipate the NREN they will rely on a few years from now? How are we preparing for (reacting to) this challenge?

SIG/ED presents a panel with diverse relationships to networking:

Vanessa Verkade, librarian at Northeastern University, speaks from the front line where she helps colleagues, faculty, and students with networked information resources. She will discuss what today's graduates need to know to master the networked universe.

Three faculty members will describe various approaches to introducing networking issues and skills: Gregory Newby, University of Illinois; Ronald Doctor, University of Alabama; and Scott Barker, University of North Carolina at Chapel Hill. Moderator: Debora Shaw, Indiana University.

The Teaching of Network Navigation Skills

Gregory B. Newby
Graduate School of Library and Information Science,
University of Illinois at Urbana-Champaign
email: gbnewby@alexia.lis.uiuc.edu

Prof. Newby will describe his experiences of teaching computer networking. From small workshops to semester classes, there are a variety of opportunities to introduce neophytes to the larger electronic world. Over only one semester, students can be brought from near computer illiteracy to networker extraordinaire: not only successfully navigating the network, but independently identifying new information resources. Newby advocates two general viewpoints for teaching about networking. First is the belief that access to information is empowerment: by learning to navigate the nets, one can expand his or her personal information resources. The second starting point is that computer networking is an emerging form of communication, for which the norms of conduct are yet emerging. New network users must learn new rules for behavior, but also have the opportunity to shape the rules as they evolve. Specific recommendations for teaching networking, including a semester course syllabus, will be discussed.

---

[1] Charles R. McClure *et al. The National Research and Education Network (NREN): Research and Policy Perspectives.* Norwood, N.J., Ablex, 1991. p. 43.

Authors Index

ASIS AND ITS MEMBERS

For over 50 years the leading professional society for information professionals, the American Society for Information Science is an association whose diverse membership continues to reflect the frontiers and horizons of the dynamic field of information science and technology. ASIS owes its stature to the cumulative contributions of its members, past and present.

ASIS counts among its membership some 4000 information specialists from such fields as computer science, management, engineering, librarianship, chemistry, linguistics, and education. As was true when the Society was founded, ASIS membership continues to lead the information profession in the search for new and better theories, techniques, and technologies to improve access to information through storage and retrieval advances. And now, as then, ASIS and its members are called upon to help determine new directions and standards for the development of information policies and practices.

Mission

The mission of the American Society for Information Science is to advance information professionals and the field of information science. This will be achieved by (1) providing methods of communication and continuing education for information professionals; (2) encouraging research, development and applications that advance the field of information science; and (3) increasing public awareness of the field and its benefits to society.

Vision

The ASIS vision is to become the foremost interdisciplinary information association in the world. To accomplish these goals, ASIS:

- Stimulates participation and interaction among its members by affording them an environment for substantive professional exchange;
- Encourages and supports personal and professional growth through opportunities for members to extend their knowledge and skills, develop and use professional networks, pursue career development goals and assume leadership roles in the Society and in the information community;
- Increases the influence of information professionals among decision-makers by focusing attention on the importance of information as a vital resource in a high-technology age;
- Promotes informed policy on national and international information issues by contributing to the formation of those policies; and
- Supports the advancement of the state of the art and practice by taking a leadership position in the advocacy of research and development in basic and applied information science.

History

The American Society for Information Science (ASIS) is a nonprofit professional association organized for scientific, literary, and educational purposes and dedicated to the creation, organization, dissemination and application of knowledge concerning information and its transfer. Founded in 1937 as the American Documentation Institute (ADI), ASIS has a history which stems from the earliest days of the modern era of documentation. ASIS today continues the spirit of the achievements compiled by ADI in its early years of existence: development of microfilm readers, cameras and services; fostering negotiations and research which resulted in the so-called "gentleman's agreement covering the photoduplication of copyrighted materials; establishment of programs for the storage and reproduction of auxiliary publications in support of journal editors; operation of an Oriental scientific literature service during World War II; support of Interlingua to foster international science communications; and cosponsoring of the 1958 International Conference on Scientific Information.

Individual Membership Application

☐ New Member

asis AMERICAN SOCIETY FOR INFORMATION SCIENCE

Please print or type in black ink.

Name (Last, First, Middle) ____________________ Day Phone( ) ____________ Ext: ______

Title ____

Organization ____

Mailing Address ____

City ____________________ State ________ Zipcode ________ Plus 4: ________

Province (Outside U.S.) ____________ Country ____________ Mail Code ____________

Check One—The Above Address Is My ☐ Work ☐ Home

Please select category of membership:

☐ Regular $85 $ ________

☐ Student $34 $ ________

I am a full-time student at ____________________

Faculty advisor's signature ____________________

Special Interest Group (SIG) Dues

☐ Check here if one free SIG selected.

Additional SIGs ________ x $6 each $ ________

Total Membership and SIG Dues $ ________

Contribution to ASIS Scholarship Fund $ ________

Contribution to ASIS Development Fund $ ________

Total Payment Enclosed $ ________(US)

Chapter Membership: You will automatically become a member of the chapter serving your geographic area (if one exists). Information about your chapter and additional chapters you may want to join will be sent to you upon receipt of your membership application.

Check or money order enclosed (Payable to ASIS)
or charge my ☐ VISA ☐ Mastercard

Account # ____________________

Expiration Date ____________________

Signature ____________________

Because your ASIS membership is individual rather than organizational, your membership goes with you if you make a career move. All membership fees and contributions are tax-deductible to the full extent of the law in the United States. You will receive notification of your active membership when your application form is processed. Please allow 4-6 weeks for delivery. For new members, membership is for one year from the month in which dues are received. Membership fees are non-refundable.

Of the annual membership dues, $11 is payment for the *Bulletin;* $15 is payment for the *Journal.*

All ASIS members may select one SIG at no charge. You may join additional SIGs at $6 each. Please note *all* selections here. Enclose payment of $6 for the second and succeeding selections.

- ☐ **Arts and Humanities** (AH)
- ☐ **Automated Language Processing** (ALP)
- ☐ **Biological and Chemical Information Systems** (BC)
- ☐ **Behavioral and Social Sciences** (BSS)
- ☐ **Classification Research** (CR)
- ☐ **Computerized Retrieval Services** (CRS)
- ☐ **Education for Information Science** (ED)
- ☐ **Foundations of Information Science** (FIS)
- ☐ **Human-Computer Interaction (HCI)**
- ☐ **Information Analysis and Evaluation** (IAE)
- ☐ **International Information Issues** (III)
- ☐ **Library Automation and Networks** (LAN)
- ☐ **Law and Information Technology** (LAW)
- ☐ **Medical Information Systems** (MED)
- ☐ **Management** (MGT)
- ☐ **Numeric Data Bases** (NDB)
- ☐ **Office Information Systems** (OIS)
- ☐ **Personal Computers** (PC)
- ☐ **Information Generation and Publishing** (PUB)
- ☐ **Storage and Retrieval Technology** (SRT)
- ☐ **Technology, Information and Society** (TIS)

Please mail this form with your payment to
American Society for Information Science
Ben Franklin Station
P.O. Box 554
Washington, DC 20044-0554

All other correspondence to ASIS Headquarters,
8720 Georgia Avenue, Suite 501, Silver Spring, MD 20910
Tel. 301-495-0900 FAX 301-495-0810